Bowling Green State University

From the SelectedWorks of Pavel Anzenbacher

May 17, 2010

A Practical Approach to Optical Cross-Reactive Sensor Arrays

Pavel Anzenbacher, Jr., *Bowling Green State University*
Premsyl Lubal
Pavel Bucek
Manuel A. Palacios
Maria E. Kozelkova

Available at: https://works.bepress.com/pavel_anzenbacher/5/

A practical approach to optical cross-reactive sensor arrays†‡

Pavel Anzenbacher, Jr.,*[a] Přemysl Lubal,*[b] Pavel Buček,[b] Manuel A. Palacios[a] and Maria E. Kozelkova[a]

Received 17th May 2010
DOI: 10.1039/b926220m

Supramolecular analytical chemistry has emerged as a new discipline at the interface of supramolecular and analytical chemistry. It focuses on analytical applications of molecular recognition and self-assembly. One of the important outcomes of the supramolecular analytical chemistry is the understanding of molecular aspects of sensor design, synthesis and binding studies of sensors while using rigorous methods of analytical chemistry as a touchstone to verify the viability of the supramolecular aspects of the sensor design. This *critical review* provides a simplified version of the chemometric procedures involved in realizing a successful analytical experiment that utilizes cross-reactive optical sensor arrays, and summarizes the current research in this field. This review also shows several examples of use of described chemometric methods for evaluation of chemosensors and sensor arrays. Thus, this review is aimed mostly at the readers who want to test their newly-developed chemosensors in cross-reactive arrays (169 references).

Introduction

A chemical sensor is defined as "a device that responds to a particular analyte, *i.e*. ion or molecule of interest, in a selective way through a physical or chemical interaction, and can be used for qualitative or quantitative determination of the analyte."[1] Most chemical sensors consist of two functional parts. The *receptor* is a moiety capable of converting the changes in the chemical composition of the molecular environment into a change in physical or chemical property, typically a change in electron distribution, energy of frontier orbitals, redox potentials, *etc*. The *transducer* is then a moiety that transforms and amplifies the perturbed properties into an observable analytical signal output whether it is optical (colour, luminescence), electrochemical or other. In most instances, the selectivity for a specific analyte originates mostly from the receptor moiety; however, transducer often modifies the output signal acting as a "filter".[2] Chemical sensors are usually divided according to a chemical or physical feature employed in the recognition mechanism and operating principle of the transducer:[1–3]

- Optical—the interaction of an analyte with the receptor subunit produces changes of optical properties. This change can be monitored and quantified by measurement of *e.g*. absorbance, luminescence, reflectance, scattering, *etc*.

[a] *Center for Photochemical Sciences and Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA. E-mail: pavel@bgsu.edu; Fax: (419) 372 9809; Tel: (419) 372 2080*
[b] *Department of Chemistry, Masaryk University, Kotlářská 2, 61137, Brno, Czech Republic. E-mail: lubal@chemi.muni.cz; Tel: (0) 420-549 49 5637*

† Part of a themed issue on the supramolecular chemistry of anionic species.

‡ This work was supported by the NSF (EXP-LA 0731153 to P.A., CHE 0750303 to P.A.), and Ministry of Education of the Czech Republic (grant ME09065 to P.L. and P.A.).

Pavel Anzenbacher, Jr.

Dr Pavel Anzenbacher, Jr. received his PhD degree in organic chemistry from the graduate school of the Czech Academy of Sciences at Prague in 1997. After completing postdoctoral research at the University of Texas at Austin with Prof. Jonathan L. Sessler, he joined the faculty at Bowling Green State University in 2000 where he is currently an associate professor. His research is focused on various aspects of materials chemistry and photophysics. His current research interests include investigations of sensor arrays, fluorescent chemosensors, organic electronics, OLEDs and OPVs.

Přemysl Lubal

Dr Přemysl Lubal received his PhD degree in analytical chemistry from the Masaryk University at Brno, Czech Republic, in 1996 under the auspice of Prof. J. Havel. After postgraduate studies at ETH Zurich with Prof. G. Anderegg and ECPM Strasbourg with Dr A.-M. Albrecht-Gary he started his independent career at the Masaryk University at Brno where he is currently an associate professor. His research is focused on the studies of coordination properties of macrocyclic ligands and application of their complexes in medicinal and analytical chemistry.

 This journal is © The Royal Society of Chemistry 2010

• Electrochemical—the analyte induces changes in electrochemical properties (*e.g.* current in a voltammetric sensor, potential in an indicator electrode—ISE, redox, *etc.*).

• Electrical—the analyte-sensor interaction results in changes of electrical properties (*e.g.* conductivity, permitivity, *etc.*).

• Mass/heat/magnetic—the analyte-sensor interaction leads to changes in mass, temperature, or magnetic properties.

While the output signal can be measured in many different ways, the simultaneous measurement of signals generated by an *organized set of responsive elements*, probes or chemosensors combine numerous advantages. For the purpose of this review, we will call such a sensor composed of an organized set of responsive elements an array sensor or shortly "an array". The sensor array literature describes various experiments arranged in high-throughput multi-well plates, microtiter plates, *etc.* Some studies have the appearance of arrays but not all such studies handle the data outputs as an array, *i.e.* the whole data set comprising the variability and multi-dimensionality of the outputs. Here, we will consider only such experiments that render the output data sets as the whole "result", which is handled, explored, and evaluated as a whole using appropriate mathematical and statistical tools. Therefore, the methods of handling the output data or signal are rather important, whilst the methods the data are processed, evaluated and interpreted are the features of distinction, rather than a tool or experimental technique. That is because the issue of sensor array methodology in analytical or supramolecular analytical chemistry is a factor of utmost importance.

In the last two decades, arrays and microarrays have become increasingly diverse tools for biological studies but their use continues to expand rapidly outside the bio-related disciplines. Likewise, the underlying array technologies, formats, and protocols continue to evolve. Traditionally, arrays have consisted of collections of distinct capture molecules such as cDNAs or oligonucleotides attached to a substrate—usually a glass slide—at predefined locations within a pattern.[4–6] Since their introduction in the 1990s,[7] microarrays have expanded rapidly into every major area of biological research, including gene expression,[8] signal transduction,[9] genome mismatch scanning,[10] inflammation,[11] cancer,[12] cell cycle,[13] DNA replication,[14] oxidative stress,[15] hormone action,[16] apoptosis,[17] neurodegenerative disease,[18] cytoskeleton,[19] and protein trafficking.[20] The capture agents used in arrays include not only DNA, but also proteins,[21] carbohydrates,[22] drug-like molecules,[23] cells,[24] tissues,[25] *etc.*

Particularly as a result of device miniaturization and micro-fabrication, whether these are electrodes, field-effect transistors, printers/spotters, or charge coupled devices (CCDs), the sensors composed of multiple elements achieved immense popularity, and were demonstrated to yield intriguing results. While most of the recent publications describe electrochemical arrays often called electronic noses or tongues,[26] this review is focused on arrays utilizing optical sensors based mostly on absorption or luminescence.

Of late, optical microarrays were applied by a number of research groups to identification and sensing of chemical entities such as ions (*e.g.* Anslyn, Anzenbacher, Severin, McDevitt, Wolfbeis, Y. T. Chang), vapors (Suslick, Walt, Lewis, Jurs), organic small neutral molecules (Suslick, Severin, Stojanovic, Anslyn, Lavigne, Shimizu, Singaram, Hamachi, Jurs), and biomacromolecules (Hamilton, Rotello, Hamachi). Selected examples of applications of sensor arrays in multicomponent analysis are given in Table 1. The topic of pattern recognition in optical sensor arrays and chemometric methods was also highlighted recently.[27]

Prior to delving into the details of optical arrays, we will discuss the term "chemosensor" or "probe" used to designate a chemical or material reacting to the presence of a chemical stimulus (*e.g.* fluorescent crown ether). Additionally, we will

Pavel Buček

Pavel Buček received his BS and MS in chemistry from the Masaryk University at Brno, Czech Republic. Pavel is currently pursuing his PhD degree with Prof. Lubal. His research interests lie in the studies of physical properties of biomacromolecules using chemometrics methods.

Manuel A. Palacios

Dr Manuel A. Palacios received his BS in chemistry from the Universidad Simon Bolivar, Venezuela, and PhD degree from the Bowling Green State University in 2008 for his work on sensor arrays under the guidance of Prof. Anzenbacher, Jr. After graduating he joined the group of Prof. David R. Walt at the Tufts University and pursues research in bioanalytical chemistry and nanoscience.

Maria E. Kozelkova

Ms Maria E. Kozelkova obtained her BS in chemistry from the St. Petersburg State University, Russia, and is currently pursuing her PhD degree with Prof. Anzenbacher, Jr. Her research interests lie in the area of protein sensing and sensor arrays.

Table 1 Examples of application sensor arrays

Analyte(s)	Sensor	Mode	Chemo. method	Ref.
Dipeptides	Cu(II), Ni(II) – AI, MCB, GCR	Abs	LDA	31
20 Natural Aminoacids	Cp*Rh – CB, XO, Gallocyanine	Abs	LDA	32
Aminosugars, Aminoglycosides	Cp*Rh – Gallocyanine	Abs	LDA	33
Tripeptides	Cu(II), Ni(II) – AI, MCB, GCR	Abs	LDA	34
Pyrophosphate, AMP,cAMP, ADP, ATP, GMP	Cp*Rh – Gallocyanine, MY, EB	Abs	LDA	145
ADP, ATP, pyrophosphate	Cp*Rh – GCR, Azophloxine, NBB	Abs = f(t)	LDA, ANN (quant.)	146
Peptide Hormones–Angiotensin I and II	Cu(II), Ni(II) – AI, XO, MCB	Abs = f(t)	LDA	35
Dipeptides (Kallidin, Bradykinin, Carnosine, *etc.*)	Cp*Rh, (*p*-cymene)*Ru,, Pd(en) – CB, MCB, NFR, NMA	Fluo = f(t)	PCA	36
AMP, ATP, GTP	Combinatorial peptide sensor array	Abs	PCA	37
Chiral vicinal diols	3 chiral boronic acid receptors and 3 pH indicators	Abs	PCA, ANN (quant.)	38
Hg(II), Pd(II), Cu(II), Fe(II), Ni(II)	5 thiols plus squaraine dye	Fluo	PCA	39
5 thiols	Hg(II), Pd(II), Cu(II), Fe(II), Ni(II) plus squaraine dye	Fluo	PCA	39
Ovalbumin, fetuin, lysozyme, BSA, elastin	Combinatorial peptide sensor array	Abs	PCA	40
Tripeptides	Combinatorial peptide sensor array	Abs	PCA	41
Trp, Phe, Leu, Val, Leucine (D-, L-form)	Cu(II) – N-donor ligands, PCV, CCR, CAS	UV	PCA	107
Chiral Amines	Cu(I)-BINAP	CD	LDA, MLP-ANN (quant.)	42
Ion Pairs (metal cations–Li, Na, K, NH_4^+, TBA^+; anions – halide, acetate, nitrate, phosphate	Polyurethane matrix + Indicators (fluorescein derivatives)	Fluo	LDA	43
10 cations (Ca, Mg, Cd, Hg, Co, Zn, Cu, Ni, Al, Ga)	9 organic analytical reagents	Fluo	LDA	103
10 cations (Ca, Mg, Cd, Hg, Co, Zn, Cu, Ni, Al, Ga)	8-Hydroxyquinoline-based sensors	Fluo	PCA, LDA	44
ADP, ATP, pyrophosphate, phosphate	Tripodal sensors	Fluo	PCA	117
10 anions (halides, phosphate, sulfate, nitrate, *etc.*)	Calix[4]pyrole derivatives	Abs	PCA, HCA	45
12 amines	24-dye sensor array	Abs	PCA, HCA	46
18 commercial beers	36-dye array (gas phase); 25-dye array (liquid phase)	Abs	PCA, HCA	57
100 volatile organic compounds	36-dye sensor array (gas phase)	Abs	PCA, HCA	47
18 commercial soft drinks	25-dye sensor array	Abs	PCA, HCA	116
14 mono- and disaccharides, artificial sweetener	16-dye nanoporous array	Abs	PCA, HCA	115,48
10 commercial coffees	25-dye sensor array	Abs	PCA, HCA	101
15 nucleotides, phosphate and pyrophosphate	Cationic polythiophene derivatives	Abs	LDA	49
Alkaloids (strychnine, brucine, cocaine, *etc.*)	Three-way junction-based sensors	Abs	KNN, SVM	160
44 metal ions	New York Tongue (47 off-the-shelf organic dye)	Abs	PCA, HCA	106
5 metal ions (Ca, Cu, Ni, Zn, Cd)	Org. reagents (FluoZin1, BTC-5N, Calcein, Lucifer Yellow, Oregon Green, Bapta-5N, Fluo-5N, *etc.*)	Fluo	SVM	50
12 D-Saccharides (ribose, arabinose, rhammose, xylose, glucose, mannose, galactose, *etc.*)	Sensor array of boronic acids with appended bipyridinium moieties	Fluo	LDA	51
Proteins (*e.g.* myoglobin, cytochrome C, ferredoxin)	16 and 8 sensor array (porphyrin derivatives with aminoacid and dipeptide pendant arms)	Fluo	PCA	113
Arylamines	Molecule-imprinted polymeric sensor array	Abs	LDA	52
Aliphatic, aromatic amines and polyamines	Polythiophene-carboxylic acid derivatives	Abs	LDA	53
Normal, cancerous and metastatic cells	Sensor array of conjugated fluorescent polymers	Fluo	LDA	54

Abbreviations used: BSA—Bovine serum albumin; GCR—Glycin cresol red; AI—Arsenazo I; XO—Xylenol orange; MCB—Methylcalcein Blue; CB—Calcein Blue; MY—Mordant Yellow; EB—Evans Blue; NBB—Naphthol Blue Black; NFR—Nuclear fast red; NMA—N-methylanthranilic acid; PCV—Pyrocatechol violet; CCR—Chromoxane cyanin R; CAS—Chrome azurole S; BINAP—2,2′-bis(diphenylphosphino)-1,1′-napthyl. Fluo—Fluorescence; Abs—Absorbance; Abs = f(t)—Absorbance kinetics; Fluo = f(t)—Fluorescence kinetics LDA—Linear Discriminant Analysis; (MLP)ANN—(Multi-Layer Perceptron) Artificial Neural Networks; PCA—Principal Component Analysis; HCA—Hierarchical Cluster Analysis; KNN—K-Nearest Neighbors; SVM—Support Vector Machines.

also use the term ''sensor element'' when referring to a single feature (well or a spot) of the sensor array, which is not further separable to lower constituents based on the response to the analyte (Fig. 1). This surely deserves an explanation. A sensor array using chromoionophore–dimethylsulfoxide (DMSO) solutions in a 96-well plate (Fig. 1A and B) will be constituted of 96 sensor elements, which are the same as the 96 wells. Based on the analytical response, each of the chromoionophore–DMSO sensor elements cannot be further easily separated to the chromoionophore and DMSO solvent because DMSO does not yield any analytical response (Fig. 1B). Thus, each well of the plate stands for one sensor element. On the other hand, multiple chemosensors, each dispersed in a polymer matrix, can be deposited on the bottom of a single well in the 96-well plate. In this case, the well is not the sensor element. Deposition of individual chemosensor-polymer materials is shown in Fig. 1D and E. In Fig. 1D, for example, each well comprises 9 sensor elements. In most arrays, however, each spot or well is one sensor element.

In general, in optical chemosensors or probes the response is a function of the analyte concentration, and one can therefore consider constructing a calibration curve, in which the slope corresponds to the sensitivity of the sensor. The limit of detection/quantitation (LOD/LOQ) is the smallest quantity of analyte that is ''significantly different'' from the blank (from statistical point of view) while the sensor is not saturated.[3,28]

 This journal is © The Royal Society of Chemistry 2010

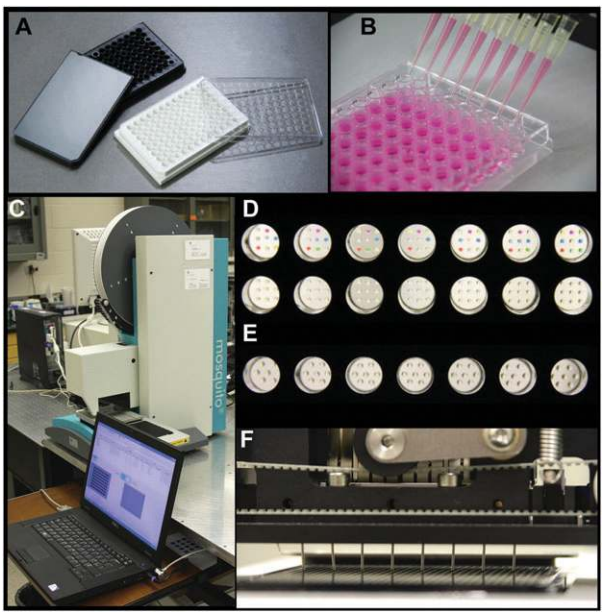

Fig. 1 A 96-well plate for colorimetric and/or fluorimetric assay (A). Sensor arrays utilizing multi-well plates are easy to prepare *via* manual dispensing of chemosensors (B) or using a robotic dispenser (C). Robotic dispensers and printers can be used to fabricate multi-element arrays within the wells: 9-member (3 × 3) array of polymer-based sensor elements deposited on the bottom of the well (D); 7 replicas of each fluorescent chemosensor-polymer membranes deposited on the bottom of the well (E). Robotic dispensers can deposit sensors into high-density (1536-well plates) for high-throughput optical-sensor arrays (F) with high precision.

Selectivity of a sensor can be defined as ability of the optical chemosensor to give response to one analyte in the presence of other analytes in the sample.[3,28–30] Most optical chemical sensors can also respond to some other analyte in a sample, *i.e.* they are not completely selective (and are cross-reactive).

The dynamic range is defined as the range of concentrations in which the sensor sensitivity is greater than zero but has not reached the level of saturation. It can be expressed as the difference between minimum and maximum analytical signal values relevant to LOD/LOQ and the highest concentration which can be used for determination of analyte concentration.[28–30]

In general, the experimental conditions (*e.g.* pH, solvent, receptor concentration, *etc.*) used for measurement with an optical chemosensor are selected to achieve minimized LOD/LOQ and other analytical parameters maximized (sensitivity, selectivity, dynamic range, *etc.*). Experimental conditions should be also adjusted to maximum robustness, *i.e.* to display a small change in analytical parameters with change of experimental conditions. However, that may not be always easy due to the environmental sensitivity of some chemosensors and detection schemes, *e.g.* impact of quenching impurities on fluorescence. Finding the best solution for these multi-factor problems is often difficult. The optical chemical sensor applied in chemical practice is validated by analyzing samples of known composition.[28–30] The precision and accuracy of an analytical procedure should be evaluated in order to estimate the source of random and systematic error.

A sensor array might be characterized by the same analytical parameters as single sensors including selectivity, detection limits, *etc.* However, meaning of these parameters may be different for the sensor array and a single chemosensor.

The principal differences between the analytical parameters of a single sensor and sensor array are:[26]

- The selectivity and detection limits of a sensor array depend not only on the physico-chemical properties of the chemosensors used to make up the array, but also on the array composition and size as well as built-in redundancy and overlaps in the parameters specific for the individual sensor elements;
- Because some aspects of the detection and discrimination limits are method related, the analyte discrimination by the sensor array can be evaluated only in case when complete experimental data set is processed using chemometric approaches and the determined parameters may depend on the calculation method adopted;
- The array sensors do not generally conform to the 1 : 1 stoichiometry (as it is in the lock and key concept) between an analyte and a number of sensor elements in the array to achieve 100% selectivity in the recognition process.

The selection of the individual chemosensors or sensor elements for an array and the sensor array size is a complex task that can be solved by different approaches. It depends largely on the knowledge of the target analyte, and whether a qualitative or quantitative analysis is sought. For qualitative purposes (classification) the dynamic range is a less important factor (as long as the analyte concentration falls within the dynamic range of most of the chemosensor elements). In the case of the limited knowledge of the analyte composition to be classified, preliminary experiments to see at least some minimal differences in the response in the color or fluorescence with naked eye (or a digital camera) is usually sufficient to pre-select the competent chemosensors. After recording the response using spectrometers or imagers the output will be analyzed using the methods described below. Additionally, there is great potential in application of information theory[55] for selecting sensors providing the highest information content.

The sensor arrays designed for application in quantitative analysis should fulfill also certain criteria. The analysis of calibration curves should give the dynamic range for a potential chemosensor considered for incorporation into the array. The calibration set used for testing sensor arrays should preferably include the mixtures of analytes of interest. The calibration curve comprising low, middle and high analyte concentrations with repetitions of measurement at those concentration levels is loaded with lower error. Therefore the experimental design should be developed to evaluate cross-reactivity of chemosensors used in the sensor array. Experimental design (ED) for a system of N analytes at two concentrations was proposed where the total number of experiments for two-level ED is 2^N, while three-level ED is increasing to 3^N experiments, *i.e.* $2^2 = 4$ experiments and $3^2 = 9$ experiments for 2 analytes. The number of measurements is increasing with the increasing number of analytes.

In order to use the experimental design for screening of the effects of chemosensors cross-reactivity in arrays, the saturated

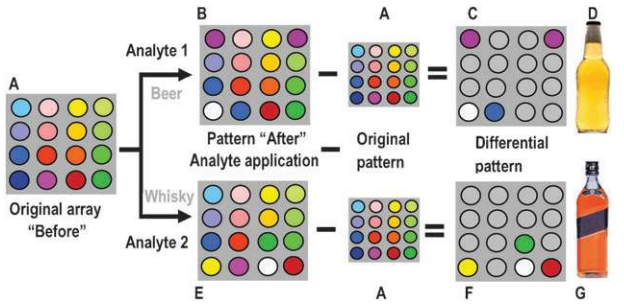

Fig. 2 Differential sensing using a hypothetical array-based sensor. To a 16-element array (A), analyte 1 (later identified as a beer) and analyte 2 (whisky) are applied. The response patterns B, E are used directly or the pattern A is subtracted to obtain differential patterns C, F. The differential patterns are identified to correspond to a beer and whisky sample, respectively.

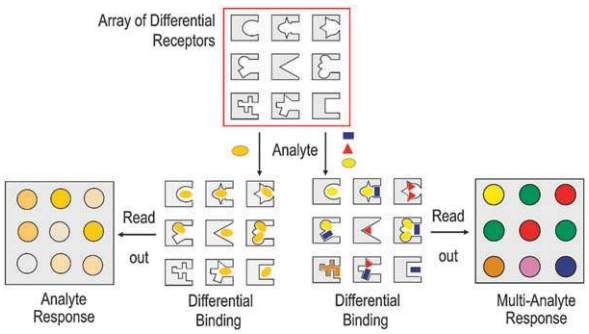

Fig. 3 Array-based sensor utilizing analyte binding by differential receptors. Regardless of whether the analyte is a single- or multi-component analyte, receptors bind a number of analytes, but each receptor binds the analytes differently thus providing a differential response.

fractional factorial designs (*e.g.* Plackett-Burman design) are frequently applied.[3,29,30] Concluding this section, the experimental design for sensor arrays should be performed. The experimental set used for testing sensor arrays should preferably include the mixtures of analytes of interest and their composition should be planned according to ED.

In the current literature, one can encounter the term "differential sensor" used in two different connotations with respect to the array-based sensing. The first, more traditional meaning is derived from the fact that the output signal from each sensor element after analyte application is directly compared to the output of the same element before analyte application and the difference, regardless of whether the output is a spectrum, reaction rate, light intensity, *etc.*, is then used in output processing, evaluation and analyte identification using pattern recognition protocols. This process is schematically explained in Fig. 2. Here, the hypothetical array sensor comprising 16 sensor elements (Fig. 2A) is exposed to two liquid samples, beer and whisky (such analyses were recently published).[56,57] As expected, the array in Fig. 2A responds to the presence of each analyte in a different way, thus producing raw response patterns 2B and 2E. These raw response patterns 2B, 2E could be used directly in a computer-assisted search for similar patterns. The term "raw response" is used here to indicate that some sensor elements in the array actually did not respond to the presence of an analyte. Alternatively, the initial pattern 2A is subtracted from the raw responses 2B and 2E to produce patterns in which only the non-zero response is shown. The resulting differential images 2C and 2F are then used to identify the analytes, beer (2D) or whisky (2G), using pattern recognition protocols.

More recently, Anslyn coined an alternative meaning for the differential sensing arrays and differential receptors (Fig. 3).[58] The term "differential" emphasizes the unifying aspect utilized in array sensors[59] that the individual receptors all interact differently and each analyte yields a different response pattern.[58] The recognition, even though it originates at the molecular level, is then observed at the level of the response patterns and the difference between them.

Protein and gene chips are therefore also differential sensor arrays. However, differential sensors may also include receptors that are cross-reactive (Fig. 3). Each receptor may bind a number of analytes, but each receptor binds the analytes differently than all other receptors. In this case, the signals of all receptors are interpreted by a pattern recognition protocol, and the result is a fingerprint-like analyte response.

In the last decade, the interest in molecular sensing has been slowly shifting from selective sensors toward sensor arrays utilizing pattern recognition.[60–62] This approach promoted the development of so-called chemical noses,[63–65] tongues,[66–68] *etc.* utilizing cross-reactive recognition sensor elements that are not particularly selective toward specific analytes. The specificity originates from recognition of the response pattern derived from an output signal (*e.g.* optical,[65,66] electrical[69,70]), comes from a large number of sensor elements. Here, a large number of sensor elements with a different degree of interactions result in the formation of a pattern specific for a given analyte mixture. Such pattern recognition approaches could be powerful enough to circumvent the need for difficult-to-make analyte-specific sensors. Also important is the fact that selective recognition tools (chemosensors, sensor elements) cannot be easily developed for a previously unknown analyte (*e.g.* yet-unidentified mutations, unexpected contaminants, previously undetected pollutants). In general, pattern recognition methods have a distinct advantage as they can show that "a response is very similar to factors (analytes) X and Y, but less similar to Z, yet is not exactly X, Y or Z". Given the number of mutations in bio-macromolecules, potential contaminants in industrial processes or pollutants and water, such information would be valuable. In the design of any assay, one should take advantage of any structural feature or ligand known to bind the likely contributors to the analyte to lower the number of cross-reactive sensor elements.

The interplay between the selectivity and cross-reactivity in sensor arrays

Conventional microarrays prefer the use of specific recognition motifs. Thus, array-based diagnostic tools are focused mostly on specific interactions such as DNA microarrays (Genechips) utilizing DNA–DNA or DNA–RNA hybridization. Similarly, protein microarrays (Biochips) utilize specific protein–substrate or antigen–antibody recognition, *etc.* usually with fluorescence detection. While these are powerful bioanalytical

This journal is © The Royal Society of Chemistry 2010

diagnostic tools, the preparation of such arrays is far from inexpensive as it requires cloning, expression or amplification of specific nucleic acid or protein sequences. Biomaterials such as nucleic acids or proteins have been endowed by the features that allow for selective recognition and sensing. This is less the case of small molecules, which generally display a lower number of recognition features, wide range of unrestricted conformations and significant structural diversity. This implies that the selective sensing of small molecules by array sensors may not be as easy as one would wish. Here, we examine if this potential lack of selectivity in array sensing of small molecules and ions is, actually, a insurmountable problem, or whether this could be one of the "life's lemons" from which we could make the proverbial lemonade.

The founding principles of supramolecular chemistry such as recognition, self-assembly, shape complementarity, preorganization[71–73] were successfully applied in numerous instances and may be used to describe association of molecules prone to form complexes. However, that presumes an insight that is not always easy to achieve or may require complex experiments (NMR and X-ray crystallography) or binding partners (ligands) that may be difficult to obtain, or are not known at the beginning of the screening. Hence, simpler materials or receptor molecules could help but are less likely to provide unambiguous answers compared to the highly complementary receptors. The approach that is capable of extracting information from a binding process involving binding partners that are not perfectly complementary may still be valuable. A receptor that lacks some of the features that generate the selectivity is likely to be a promiscuous one. The upside of cross-reactive receptors is an easier preparation and, sometimes, faster dissociation kinetics. The downside is that an assay based on cross-reactive receptors may require a larger number of recognition elements, and may result in a relatively small magnitude of response, which may complicate data reading, and evaluating. The number of the necessary recognition elements can be lowered by selection of the elements that show the best recognition capability (depending on an application, this could mean affinity, selectivity, LOD, *etc.*), or by upgrading the selection process by attaching a specific ligand (*e.g.* biotin to increase the selectivity for streptavidin). In the assay design, it would be foolish not to take advantage of known selective binders for key analytes[74] (if such structural features are known) to lower the actual number of cross-reactive sensor elements required for recognition. Finally, the presence of unknown analyte components could be much easier to detect by less specific cross-reactive elements, which leave an opportunity open for the unknown.

Selective *vs.* cross-reactive sensor arrays; is there a middle ground?

In sensor arrays containing differential chemosensors, the sensing elements often display a certain degree of selectivity towards certain analytes, whether this partial selectivity is due to receptor-analyte binding or other interactions affecting the signal transduction (*e.g.* heavy atom fluorescence quenching). The potential problem with highly cross-reactive sensing elements is that the information content that each individual sensor generates is often low and only the simultaneous analysis of a number of such sensor elements generates a differential response pattern with enough information. Increasing the number of read-out channels (multiplexing) often leads to increase in the amount of available, and the sensing process results in an accurate analysis.

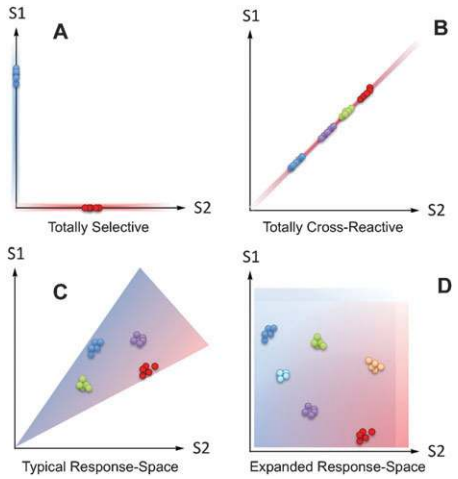

Fig. 4 Schematic representation of the response-space of two sensors S1 and S2. (A) S1 and S2 are 100% selective. (B) S1 and S2 are 100% cross-reactive to a group of analytes. (C) S1 and S2 are cross-reactive, but have certain selectivity towards some analytes. (D) S1 and S2 are cross-reactive and have enhanced selectivity towards key analytes, hence expanding the potential response-space.

To illustrate the issues associated with selectivity and cross-reactivity in array responses, Fig. 4 shows a hypothetical response space generated by a sensor array of two hypothetical sensors, S1 and S2. First, in the case that S1 and S2 are 100% selective towards one analyte each, space resolution is achieved, but the sensor array can respond only to these two analytes. In the second case, S1 and S2 are 100% cross-reactive sensors. Here, several analytes can be detected, but the space resolution is lower, to a point where any additional sensors provide almost the same information as the two sensors, and the lack of resolution makes extracting of useful information difficult. The third case is typical for most cross-reactive supramolecular sensor arrays. The cross-reactivity of sensors S1 and S2 leads to an expansion of the space response and improved resolution might be achieved. Also additional analytes could fall within the response space. Finally, the fourth case presents the middle ground between the first and the third case. The response-space is expanded when S1 and S2 are known to be cross-reactive in their response toward certain analytes whilst their selectivity is biased toward some key components. As a consequence, the resolution of the sensor array could be increased.

The selectivity in the sensing elements could potentially be the key component to increase the information density generated by the sensor array. Hence, the discriminatory power might be improved by the use of selective yet cross-reactive chemosensors. The increase in the discriminatory capability (orthogonality) of the sensor elements should allow a significant reduction of the number of array elements,

while maintaining the reliability of the sensing process. Perhaps, because such a potential reduction of sensor elements can be intuitively derived, there is no explicit report of this approach in the literature, but it has become one of the cornerstones in the design of our high-resolution minimal-size (a number of sensor elements) sensor arrays (see below).

Considering that the response-space alluded to above may be constructed using almost an unlimited number of observed variables (absorbance, emission intensity, red-ox potential, impedance, *etc.*), it is clear that such response spaces are very much multi-dimensional in their nature, and that the response of the array sensor may be distributed among these multiple dimensions. Not surprisingly, analyses of the multi-dimensional responses require somewhat sophisticated analytical tools, specifically pattern recognition protocols capable of reducing the dimensionality of the response space to the dimensions carrying majority of the response information and finding the significant response patterns.

Optical arrays: absorption and light emission intensity as a mode of signal transduction

A vast number of array sensors in bio-application and analytical chemistry utilize changes in optical properties of the sensor elements to provide output signal.

Typically, these observed optical properties are change in color (absorption spectra) and luminescence emission intensity. Needless to say, there are also array sensors that utilize other modes of signal transduction such as luminescence lifetime,[75] conductivity,[76,77] potentiometry,[78,79] impedance,[80,81] radiofrequency,[82] or polarographic[83] methods.

Here, we will discuss some of the methods used to obtain the signal from optical arrays. In the optical sensor arrays, it is desirable to record the response from all sensing elements on a given platform simultaneously (multiplexing).[84–87] While spectroscopic techniques are highly advantageous due to the high density of information that can be obtained,[88] they also require certain time for the data collection from each sensing element and also complex equipment (spectrometers, scanning setups) to perform the reading. In this regard, imaging techniques have the advantage of being capable of detecting several features in a given field of view. The fast development of sensitive CCD or CMOS cameras has promoted the use of imaging techniques for quantitation of fluorescence intensity. For both the spectral and imaging methods, multiple responses produced by sensor arrays require multivariate analysis for the handling and interpretation of the response patterns. For the detection and evaluation of the response data, two main approaches to instrumentation and methodology can be used: First, the spectroscopic method, which is usually based on collecting spectra, either independently (*e.g.* from a number of cuvettes) or using a spectrometer equipped with a so-called sipper and a flow-through cell (Fig. 5). Similarly, some multi-well plate readers also have the capability to record the whole spectra from each well. When the spectrum is obtained, the response data from certain wavelengths are then used as input.

Image recording from multi-well plates or array of vials with analyzed solutions using various flatbed scanners achieve a similar effect. Suslick and coworkers first introduced the flatbed scanner as an analytical tool for the detection of color intensities in sensor arrays.[87] The general idea is that once a reliable assay is developed, the sensor array could be deployed and read on any PC. In this context, other groups have recently taken advantage of widely available tools to carry out chemical analysis. Various "consumer" technologies such as cell phone cameras[89] by Whitesides and DVD readers by Potyrailo,[90] were implemented for the optical detection of changes in sensing elements. The obvious difference is that the spectrometers yield more finely split output channels while the common flatbed scanners usually provide only three or

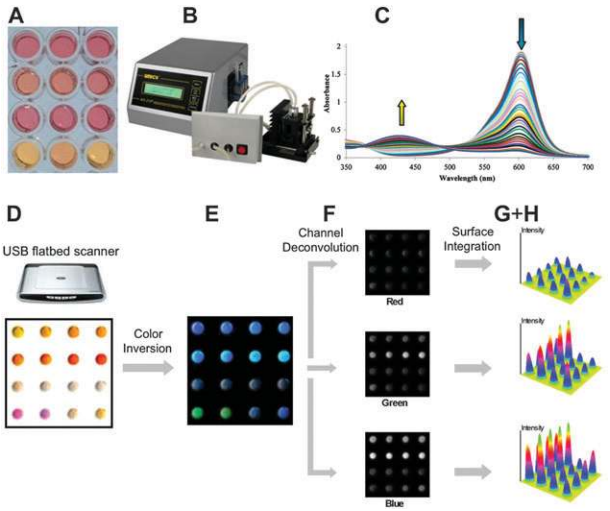

Fig. 5 Colorimetric array sensing. The solutions from the wells (A) in multi-well plates are analyzed by UV-vis spectrometer, usually equipped with a flow-through cell and a liquid handler (sipper) (B). The changes in absorption spectra (C) are then analyzed. Alternatively, a colorimetric array may be processed as an image using a scanner as follows: Image recording-scanning (D), color (RGB) inversion (E), RGB deconvolution (F), averaging (G) and integration (H) of the gray value of the non-zero pixel for each sensing element.

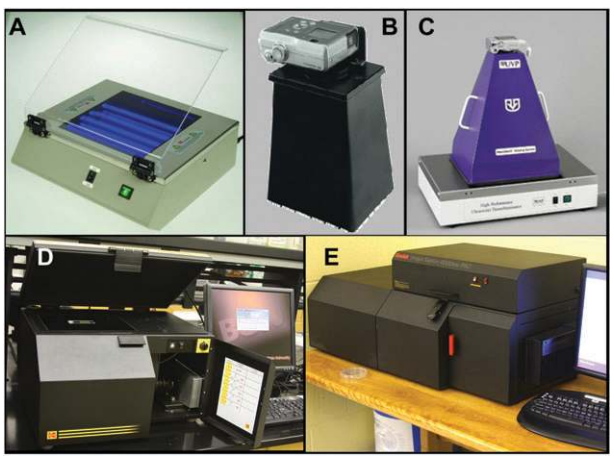

Fig. 6 Array imaging can be performed using UV-trans-illuminator (A) and a shielded digital camera (B); the whole setup (C) is available from commercial sources. Alternatively, UV-image scanning and processing is conveniently performed using professional imagers (D and E).

This journal is © The Royal Society of Chemistry 2010

four channels, usually red, green and blue (RGB), sometimes also red, yellow, green and blue (RYGB).

The general approach for the analysis of the color intensity of the image consists of: (A) recording (scanning) the image, (B) color (RGB) inversion, so the background is black (has a zero-value), (C) RGB deconvolution, (D) averaging, and finally (E) integration of the gray pixel value for each sensing element (well or spot) in the array for each (RGB) component (Fig. 5).

Reading the fluorescence-based arrays is somewhat more involved.

Most biochemistry and biology laboratories own so-called gel-documentation systems consisting of a trans-illuminator and a digital camera to capture the gel images. These can be made in house or purchased in various versions ranging from manually operated ones that utilize simple inexpensive digital cameras, to fully automatic ones with a number of features to eliminate background and increase contrast. The most important features for the in-house built systems is that the camera can focus on relatively close objects (20–50 cm), and allows for installation of a background-reducing cut-off filter.

A set of filters with various transmission wavelength region may allow to split the emitted light into various output channels. Finally, band-pass filters are the best option as they allow selecting individual colors, but are often more expensive. Perhaps the best option are epi-fluorescence image-recording systems, frequently used in biology labs for recording emission-based images from gels to live animals. While these are often expensive, they come with wavelength-selectable sources of excitation and emission wavelengths, and are often equipped with high-sensitivity cooled CCD cameras for high contrast images and software that allows selecting and evaluating the light intensity in regions-of-interest (ROIs), and are, generally, very user-friendly.

Overview of chemometric analysis of optical array sensors

Chemometric analysis of the array involves several steps, largely mathematical, aimed at pattern recognition and analysis. In general, following the data acquisition from the optical sensor array, various steps known as data preprocessing are performed. Such steps involve dimensionality reduction, exploratory data analysis (EDA), classification and clustering analysis, and finally generating a model describing the output data. The model is then verified using various analytical methods to finalize the description of the array-analyte behavior. This idea is described in the following paragraphs and in the scheme in Fig. 7.

Response data matrix and preprocessing of the data

The analysis of the data obtained from multi-sensor arrays is a very important aspect of the array sensing, one of high importance, and a topic of numerous papers and reviews.[91,92]

The recorded data are usually arranged in a response matrix containing a number of columns corresponding to the number of features detected, and a number of rows corresponding to the number of recorded observation (Fig. 8). An extra column

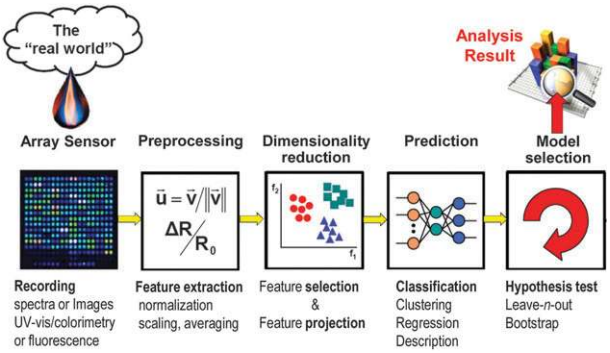

Fig. 7 Outline of the chemometric analysis.

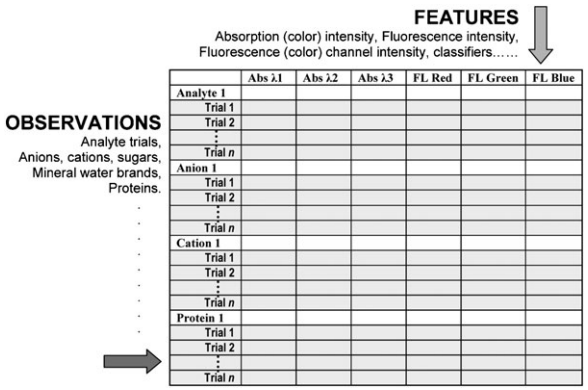

Fig. 8 Schematic representation of the data matrix.

with the classifier descriptor can be added for listing of classifier description.

Raw experimental data have an intrinsic error, *e.g.* due to systematic error or deviation due to (im)precision of measurement or instrument. In general, an important step in the data analysis is a detection of outliers (in a statistical sense), which should be excluded from the experimental data prior to pretreatment. The detection of outliers, *i.e.* responses that do not follow the same model as the rest of the experimental data, can be carried out by Dean-Dixon test for smaller or by Student *t*-test for larger data sets. This procedure is described in many analytical chemistry textbooks.[3,29,30]

While in many instances one can use raw, *i.e.* unprocessed, data, the data pre-treatment after collection should be carefully considered with understanding of what each preprocessing method means for the data set. Needless to say, several preprocessing methods should be tested prior to application of advanced chemometric techniques since the resulting responses may be significantly altered. This may be particularly important for novices in the field because in the final result of combined effects of preprocessing and, for example, artificial neural network analysis, it may not be obvious which features are due to preprocessing and which are due to the analytical method selected.

In very general terms, experimental data after testing for outliers can be preprocessed using a number of simple methods, each developed to obtain different types of information from the subsequent multivariate analysis or to simply clean up the data and achieve better resolution.

The preprocessing of the data may have a significant impact on the final outcome of the multivariate analysis. Unfortunately, the power of the data preprocessing is usually underestimated due to the lack of a general rule or procedure to guide in selecting the preprocessing methods suitable for a given application. In most cases, the preprocessing algorithms can be systematically tested before the actual multivariate analysis method is applied to achieve the desired result.

Several methods of data preprocessing have been proposed in the literature[93–96] including relative scaling, background subtraction, signal average, linearization, mean-centering, auto-scaling and range-scaling.[92] The selection of a preprocessing method to generate the response matrix requires a certain experience or systematic testing to arrive at the best method for a given data set.

For example, normalization techniques (*e.g.* relative scaling, background/baseline subtraction, signal averaging, linearization, mean-center, autoscaling/range scaling, *etc.*) are employed usually in qualitative applications since some of the relative scaling procedures can result in elimination of the concentration dependence of the sensor response intensity.[95] Therefore, it seems that it is less suitable for quantitative prediction. Another simple method is signal averaging of replicate measurement while the signal-to-noise ratio (S/N ratio) of a sample response is increased by $\sqrt{N}$ factor, where N is the number of measurement replicas. That means that 9 replicas lead to S/N $\approx$ 3, while to increase the S/N ratio twice to S/N $\approx$ 6, 36 replicas are needed. The analytical signal higher than 3-times value of noise corresponds to the sensor detection limit, thus the signal averaging can lead to decrease of LOD/LOQ. The replication of measurements in sensor arrays should be the compromise between the size of arrays and the amount of generated information. This parameter can be calculated and used for evaluation of the designed sensor arrays. Limiting the number of replicas is less important as a time-saving measure because the time needed for preparation and reading of the arrays is the less and less limiting factor now due to automation of the whole procedure.

The linearization of analytical signal can be used to assess the dynamic range of an optical chemosensor having nonlinear response, for example due to fluorescence self-quenching. Linear techniques can then be safely used for the linearized data, or alternatively, only the linear region of the concentration response curve of the sensor should be used from the whole experimental data. The following paragraph gives several examples of preprocessing methods. These are not exhaustive and additional methods can be found in the literature.[95,97]

Quite often, in the colorimetric assays, a relative baseline subtraction is applied to the data set by subtracting the color intensities in the specific color channels (*cc*), for example RGB, from the sensing elements before (C_0) and after (C) the exposure to the analyte (baseline subtraction):

$$R_{cc} = C - C_0. \quad (1)$$

Similarly, in the fluorimetric assays, a combination of the relative baseline subtraction and relative scaling is used. The relative intensity is calculated from the fluorescence intensities in the individual color channels (*e.g.* RGB, RGBY, *etc.*) subtracting the intensity from the sensing elements before (F_0) and after (F) exposure to the analyte and dividing by the signal intensity recorded before (F_0) exposure to the analyte (eqn (2)). This method is often used in qualitative applications.

$$R_{cc} = \frac{F - F_0}{F_0} = \frac{F}{F_0} - 1 \quad (2)$$

Also, in the realm of supramolecular chemistry or supramolecular analytical chemistry, where the spectroscopic experiments are often run prior to the actual assay to pre-determine the stoichiometry and relative chemosensor-analyte affinity, the corresponding affinity constant and the magnitude of response, one can conveniently use the range-scaling method. In the range scaling, the response after background subtraction is compared with the dynamic range of the response for a given analyte and a given sensor.

$$R_{cc} = \frac{F - F_0}{F_{\max} - F_0} \quad (3)$$

Even though the range scaling has yielded interesting results in some applications, in the case of cross-reactive sensor arrays its implementation may not be easy since the determination of $F_{\max}$ is relative to the dynamic range of each analyte, which is not always possible in multianalyte environments. Nevertheless, in this chapter we have included, as an example, data that have been preprocessed by both the relative intensity and range scaling to illustrate the impact the preprocessing methods have on the final outcome of the multivariate analyses.

Also, other methods may be used. In fact, some groups develop their own pre-treatment methods to suit their unique analytical situations. The authors of this review used, among others, an absolute value of range-scaling or signal normalization. The absolute value of range-scaling is described as:

$$R = \left| \frac{F - F_0}{F_{\max} - F_0} \right| \quad (4)$$

or normalization of the output signal:

$$R = \frac{F}{F_0} \quad (5)$$

The relative scaling should also be used with caution as it might lead to "relativization" of the signal output and a corresponding loss of quantitative information (see above). This means that during this preprocessing even the output channels showing a very small change in response might be scaled up to the level of stronger, and perhaps more important, signal from other sensor elements. Another potential pitfall of normalization as a pretreatment method is the possibility of increasing the noise level (decreasing the signal to noise ratio) by amplifying the weak and noisier variables.

The above list is far from complete, and other methods of feature extraction are known from the literature[95] to include various approaches to signal averaging and autoscaling (sets the mean at the origin and the variance within the data to 1), which is usually used when responses are on different magnitude scales.[92] It is most advisable to test several of the

 This journal is © The Royal Society of Chemistry 2010

feature extraction/preprocessing methods and test the clustering and/or classification. Comparison of the pre-treated and raw data in the exploratory data analysis will provide a valuable insight as to what can be expected from the data set.

Pattern recognition protocols: multivariate analysis in chemical sensor arrays

Pattern recognition techniques have been widely used for the interpretation of multivariate data sets in chemistry.[93] Notably, the intrinsic multivariate nature of the responses associated with sensor arrays requires the implementation of such methods for the understanding and evaluation of the quality of the data.[95] These methods are often based on the reinterpretation of the data into a lower dimensional space (dimensionality reduction) or on the comparison of the measurements of the vector responses direction and magnitude.[98]

Exploratory data analysis (EDA) can be seen as a contrast to traditional hypothesis testing. While the testing of hypotheses always requires a prior assumption (hypothesis) about the data (*e.g.* "There is a difference in the fluorescence upon addition of analyte A *vs.* analyte B"), the exploratory data analysis is not based on any prior assumptions. Therefore, the EDA is the important first step in any array analysis. In a typical exploratory data analysis all variables are taken into account using both graphical (*e.g.* drawing simple scatter plots) and formal methods (*e.g.* PCA—principal component analysis) to search for systematic patterns.

EDA consists mainly of the techniques of PCA and factor analysis (FA). These analyses seek to explore the data to identify the trends and clustering behavior that could reflect a large number of variables, and may not be obvious. For obvious reasons EDA methods are predominantly applied to raw data, *i.e.* prior to any pretreatment, which could amplify or reduce the magnitude of the observed features.

Often, the practical purpose of PCA in the EDA is to identify the features (variables) that are most important for the description of the chemical behavior of the sensor. Such important features could be leveraged for other analytical tasks, for example, improving signal to noise ratio and increasing the reliability of classification, reduction of the data set for simpler analysis, removing sensor elements that contribute more noise, *etc.* Such exploratory work is frequently performed using PCA, see below.

Response patterns can be analyzed by unsupervised and supervised multivariate analysis methods. The unsupervised methods, sometimes called clustering analyses, are more formal methods of treating samples that allow for detecting, and visualization similarities in the response data. The unsupervised methods do not operate with additional information, *e.g.* the sample identity. Unsupervised methods are often used to explore the actual clustering and dispersion of the data.

Factorial methods are generally used to reduce dimensionality, *i.e.* project the original experimental data set from high dimensional space onto a line, a plane, or a 3D-coordinate system using certain mathematical procedures, for example PCA, hierarchical clustering analysis (HCA), FA, singular value decomposition (SVD), eigenvector projection and rank annihilation, and others.[3,94] This is very important in sensor arrays as these often produce a high amount of experimental data in very short time. The factorial methods can help to extract the information most useful in solving the analytical problem while reducing the multiple dimensions to few (often two or three) that are easy to visualize in the form of 2D and 3D graphs, which are easy to comprehend and relay the salient information.

Not all methods, however, have been used to solve the problems of supramolecular and supramolecular analytical chemistry. In fact, the methods most frequently used for unsupervised analysis of multivariate data sets are based on clustering analyses (*e.g.* HCA), and statistical analyses (*e.g.* PCA). The supervised methods operate with the identity of the sample in the training data sets to generate models that can later be used for classification of unknowns, as it is in the case of discriminant analysis (DA). Several other multivariate analysis methods based on neural networks, data mining and machine learning algorithms have been also used in the evaluation of sensor arrays. Here, the methods used in the field of supramolecular analytical chemistry (Table 1) will be briefly explained. The readers are encouraged to read in detail the corresponding chapters in the books listed above.

Principal component analysis

Generally speaking, PCA is a statistical treatment that consists of the reinterpretation of a multi-dimensional data set into a new data set with reduced dimensionality in such a way that the most significant characteristic of the data (the variance) originally contained in the data set is preserved. This is achieved by calculating orthogonal eigenvectors (principal components, PCs) that lie in the direction of the maximum variance within that data set. The first PC comprises the highest degree of variance, and other PCs follow in the order of the decreasing variance. Thus, the PCA concentrates the information within the data set into a lower dimensional space and ranks the new dimensions in the order of their importance.

The PCA typically departs from the covariance (or correlation) matrix $\boldsymbol{C}$ of the original preprocessed data set (see above). The correlation matrix is used when the variables are heterogeneous or use different scales. PCA starting from the covariance or correlation matrix having homogenous scales usually yields similar results.

$$C = \begin{pmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m,1} & c_{m,2} & \cdots & c_{m,n} \end{pmatrix} \tag{6}$$

where

$$c_{m,n} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{im} - \bar{x}_m)(x_{in} - \bar{x}_n)$$

The premise is that the projection of the original n-dimensional data set into the reduced space can be carried out by decomposing $\boldsymbol{C}$ into a score matrix $\boldsymbol{S}$ and a loading matrix $\boldsymbol{L}$, so $\boldsymbol{C} = \boldsymbol{S}\cdot\boldsymbol{L}$. Several methods could be used to decompose $\boldsymbol{C}$.[99] Among such methods, the singular value

decomposition (SVD) is the algorithm implemented by most commercial software. This is because SVD has proven to be very robust in a number of applications.[100]

When SVD is applied directly to the covariance matrix $\boldsymbol{C}_{n,m}$ (m trials, n response features), it results in factorization of the form:

$$C = \begin{pmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m,1} & c_{m,2} & \cdots & c_{m,n} \end{pmatrix} = U \cdot \Sigma \cdot V^T. \qquad (7)$$

such that $CC^T = U$ and $C^TC = V$, where $\boldsymbol{U}_{m,n}$ and $\boldsymbol{V}_{n,n}$ (score and loading matrices, respectively) are orthonormal and formed by eigenvectors termed principal components PC_i representing the projection of C into the new principal components eigenspace. The matrix Σ (singular value matrix) is a diagonal matrix containing singular values σ_n, which are associated with the root-square of the eigenvalues ($\sigma_n^2 = \lambda_n$) for each PC_i. This implies that the minimum variance associated with the original variable is set to 1 and the sum of all j eigenvalues is equal to n of the number of variables originally comprised in C,

$$\sum_{j=1}^{n} \lambda_j = n.$$

The variance contribution of each eigenvector PC_i^{var} is determined by the portion of the λ_i divided by the sum of all j eigenvalues. The first PC contains the highest degree of variance, and the other PCs follow in the order of decreasing variance.

$$PC_i^{\text{var}} = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} \qquad (8)$$

The columns of the score matrix $\boldsymbol{U}_{m,n}$ can be used to project the covariance matrix C into a lower dimensional space for exploration of patterns in the data (clustering). This representation is called score plot and is usually associated with a certain amount of variance represented by the PC_i^{var} for each PC used in the plot.

The columns of the loading matrix $\boldsymbol{V}_{n,n}$ are composed of n vectors associated with each of the variables (or response features). When the correlation matrix is used instead of the covariance matrix, the loadings are nothing more than the correlation coefficients between the original variables and the newly derived principal components PC_n. Moreover, the loading vectors are the projection of each feature n into the principal components eigenspace. Since these vectors are orthonormal, the relative contribution of the k-th variable to the principal component PC_n is taken as the square of the loadings (correlation coefficient, r_{PC_nk}) between PC_n and the k-th variable (feature).[100]

$$k_{\text{Contribution}} PC_n = r_{PC_nk}^2 \qquad (9)$$

$$\sum_{k=1}^{n} r_{PC_nk}^2 = 1 \qquad (10)$$

The PCA yields a lot of useful information when applied to the characterization of the response of sensor arrays. For instance, from the mathematical point of view, a successful PCA is one, in which the dimensionality reduction is maximized and a number of features are compressed down to two or three of the PCs. On the other hand, sensor arrays with a high discriminatory capacity will generate responses that will be so scattered in the n-dimensional space that the PCA fails to describe most of the variance in the first two PCs. In these situations, statistically relevant PCs could be utilized to explore the patterns in the data. Several methods have been proposed to determine the numbers of PCs statistically relevant to the description of a data set.[93–95,100] Here, we use Kaiser's rule, which states that only the PCs with associated eigenvalues $\lambda_i > 1$ are relevant, since the minimum variance for a single variable was originally set to 1.

Thus, from a sensor array analysis point of view a successful PCA is one, in which the score plot shows clear clustering of similar samples and the eigenanalysis shows the total variance (information), generated by the sensor array, widely dispersed over several PCs. Such results would attest to a high discriminatory power of the array. Unfortunately, having several statistically relevant PCs often implies that even a 3D representation might not be an accurate description of the sensor array response, and it would be quite difficult to correlate response features with distances in the 2D or 3D score plots. This is the downside of the PCA. Hence, it is very common to complement the PCA with another multivariate methods such as HCA.[101,102]

Perhaps the best approach is to use PCA in those cases when the majority of the individual chemosensors (sensor elements) in the array display linear response. If this condition is not satisfied, *i.e.* the sensor elements respond in a nonlinear or nonadditive fashion, class separation and classification using PCA may not be possible. PCA is mostly used as a tool for qualitative studies,[95] however, in the cases of a linear signal output, a semi-quantitative information can also be obtained.[106]

As mentioned before, preprocessing the data might impact the final outcome of a PCA. In order to illustrate this point we took a set of data from one of our previous papers[103] and applied a preprocessing routine(s) utilizing first the relative change in the fluorescence intensity and second the range-scaling (see above). In the case of this data set it is clear how the two preprocessing procedures yield very different score plots (Fig. 9). For this data set, better clustering is obtained with range scaling (Fig. 9, bottom). However, this is not necessarily always the case and judicious choice of preprocessing must be applied to obtain best clustering and resolution.

PCA is one of the most frequently used methods in pattern recognition in the field of optical array sensors. Not surprisingly, it found application in sensing of a number of analytes including cations,[104–106] anions,[102] small molecules such as aminoacids,[107] saccharides,[108,109] explosives,[110] poisonous gases,[111] peptides[112] and proteins,[113,114] sweeteners,[115] beverages,[57,104,116] consumer products such as toothpaste,[102] *etc.* (Table 1 shows selected examples).

As an example we show the analysis of phosphates in blood serum by Anzenbacher *et al.* that employs an array of six

This journal is © The Royal Society of Chemistry 2010

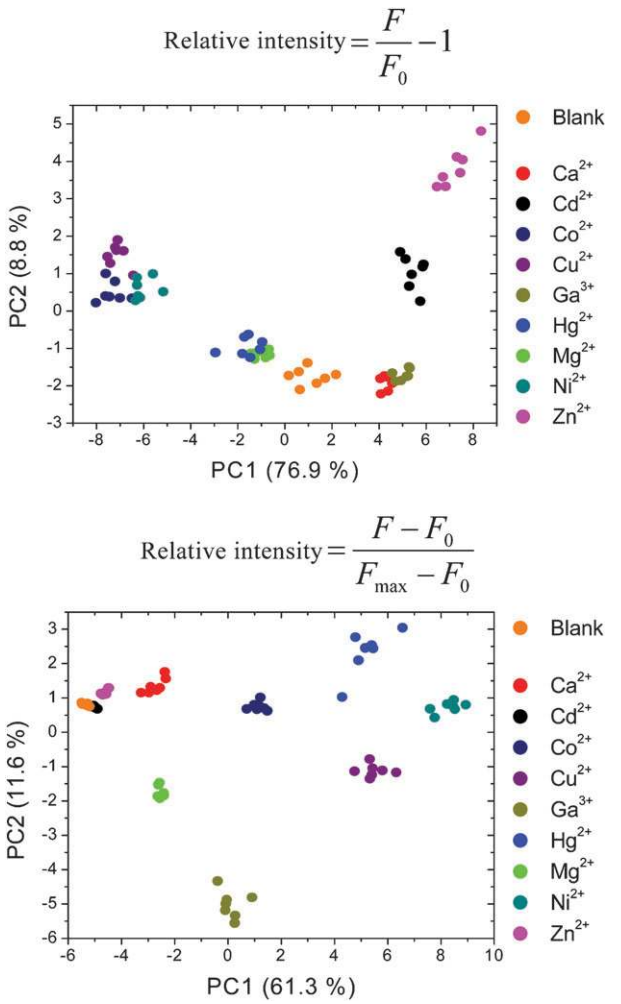

Fig. 9 Impact of the preprocessing of a data set on the PCA. The data set is composed of the responses generated by the sensor array described in the literature[103] and describes array sensor response to 9 different cations (250 μM) at pH 5. Top: Data preprocessed using relative change in the fluorescence intensity. Bottom: Data preprocessed using a range scaling algorithm.

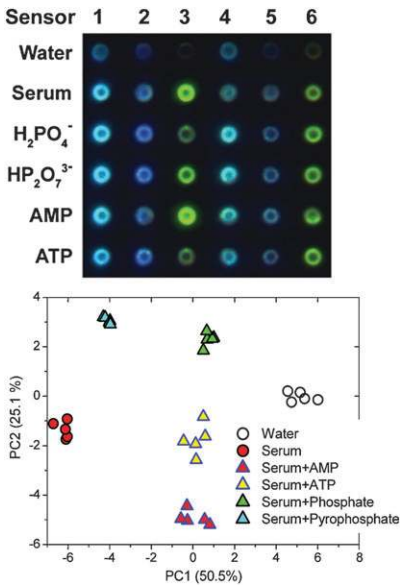

Fig. 10 Top: Qualitative changes in fluorescence of the sensor-polyurethane films after addition of human blood serum and serum with added anions (to increase the concentration to 5 mM). Bottom: PCA of the array response shows the quantified changes induced by the anions added to the blood serum, and compares the effect of anion to the pure serum and water. (This figure was reproduced with permission from Wiley-VCH Verlag GmbH & Co. KGaA, see ref. 117.)

sensors with *turn-on* fluorescence signaling.[117] It is important to note that the blood serum was not purified and contained all ions, proteins, color background, *etc*. The blood serum behaves as a unique buffer, containing also phosphates and carboxylates, and therefore gives a unique response (Fig. 10).

The blood serum with added anions (phosphate, pyrophosphate, AMP, or ATP) were analyzed. As expected, due to its intrinsic anion content, the serum itself turned on fluorescence of the array creating a unique fluorescence response, further modulated by the added anions (Fig. 10, top). PCA was then used to show that the sensor array is capable of generating distinguishable response patterns between phosphate, pyrophosphate, AMP, and ATP, as shown on the PCA score plot in Fig. 10 (bottom).

As stated above, PCA and other exploratory tools may be successfully used to study the response in array sensors, and also to optimize the sensors and the variables recorded to achieve the best resolution by the array sensor. For example, PCA was used to minimize the number of sensor elements required to achieve analyte classification, select the variables contributing most to the successful analysis, *etc*. This utility of PCA is demonstrated on the optimization of the number of sensor elements in the array sensor.

An important problem during the design of the array sensor is the question of how many sensor elements and variables are required to achieve a certain level of discrimination. Unfortunately, without testing and a training set it is hard to predict a sensor array behavior. A typical approach is to overestimate the number of sensor elements by including more sensors than is presumably necessary (entire combinatorial libraries) and pre-screen the responses for the best set of sensors.[74]

On the other hand, it has been demonstrated that the observations focused on chemosensor-analyte binding using classical spectroscopic titration experiments (but also calorimetry, voltammetry, *etc*.) in solution can be extrapolated to the behavior of chemosensors in array sensors. This is because, for example, association constants determining the affinity and selectivity of chemosensors for analytes of interest allow for a design of an array sensor likely to respond to the analytes that displayed reasonable affinity in such solution-based tests. This allows for the design of arrays biased rationally towards certain desirable analytes. Depending on the discriminatory power of the individual sensing elements in the array, it is possible to design sensor arrays with a reduced number of sensor elements still capable of identifying and quantifying certain group of analytes with high accuracy. A procedure to determine the number of sensors necessary to "resolve" an analyte set is highly desirable since it could

result in less complex and "personalized" platforms developed on-demand for a number of applications.

For any sensor array comprising a larger number of sensing elements than might be needed for an accurate classification of a given analyte, a mathematical method could be developed to reveal which sensor elements have correlated responses and thus are redundant. Conversely, such a method should also be able to identify the sensor elements that contribute most to the successful analysis. In the case of two sensors showing similar behavior in the *n*-dimensional space the contribution to the dispersion of the data of both sensors will be virtually the same. One could think about the extreme case where the same sensor is used twice in the same array. The outputs yielded by both sensors should be highly correlated thus no additional information would be obtained. For instance, one approach how to realize the redundancy of sensor elements or variables could be to search for commonalities in the HCA. Unfortunately, the HCA just yields similarities (or differences) between variables, and most sensor elements display more than one variable (response channel). Hence, just studying the similarities of each individual variable makes it difficult to decide, which sensor to remove from the array. This is because the other channels (variables) associated with the same sensor, could still display unique features. Finally, while the HCA could be used to "rectify" the data by eliminating the variables displaying contribution of low significance, it is not convenient for the systematical elimination of sensors from the arrays.

Since the response of sensor elements are *c*-dimensional, considering *c* as the number of channels used for the detection, one needs a tool that interprets the vectorial response in the *c*-dimensional space (sensor space) and correlates it with the *n*-dimensional response space generated by the entire array. As mentioned before, the PCA can describe the variance (information) of a data set in a reduced space generated by the principal components. The contribution of each variable to each principal component is given by the projection of the original variable to the principal component. Hence, the contribution of a sensor *S* (or variable) to a principal component *p* is the sum of the square of the loadings ($r^2_{cPC_p}$) of the variables corresponding to a channel *c* as shown in Fig. 11.

As has been reported in the case of mono-dimensional sensors,[118,119] the idea is to select the sensors with the maximum contribution to each of the PCs of statistical significance. Thus, after the PCA is applied to the original data set, a sensor array with *m*-elements will be reduced to a maximum number of sensors equal to the number PCs with

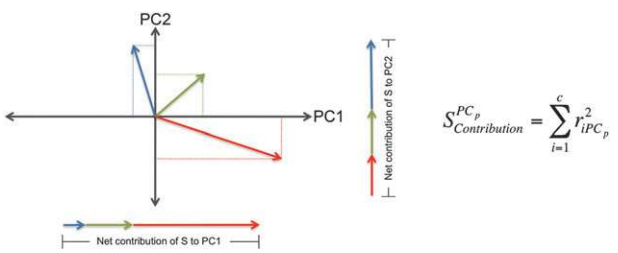

Fig. 11 Schematic representation of two not-normalized principal components (↑) showing the projections (——) of the centered RGB responses (↑↑↑) of a sensor S to the principal components.

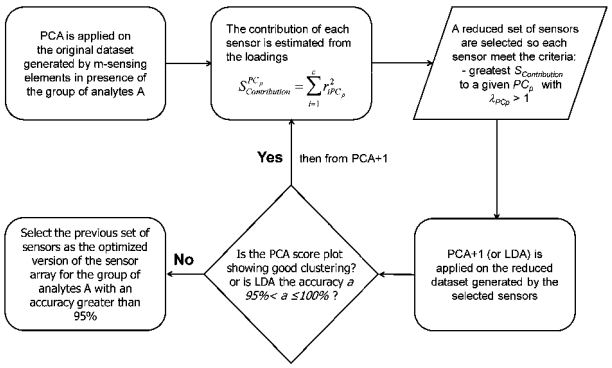

Fig. 12 Flowchart of the procedure for the sensor array size optimization. In this case the algorithm will yield a sensor array with accuracy greater than 95%, but the accuracy threshold can be adjusted to a desired level.

eigenvalues greater than one (according to Kaiser's rule). A second PCA is then carried out on the reduced version of the data set to study the amount of information lost in the previous reduction step. At this time, it is possible to employ another multivariate technique such as LDA to quantify predictability of such secondary array. If the predictability is still acceptable, the contribution of the sensor can be estimated and another reduction step can be performed. The process can be repeated until a low threshold for the predictability (or clustering in the score plot) is reached (Fig. 12).

Finally, another method based on an iterative algorithm was also proposed in the literature.[120] While iterative methods are convenient since they study all possible combinations of sensors in a sensor array, depending on the number of sensor elements it could be tedious and the additional information gained through a systematic reduction process, such as which sensor elements are shaping the space response, is not obtained as a direct consequence of the analysis.

Palacios and co-workers[104] used the previously described method to optimize the number of sensing elements in a six-member sensor array for metal-ion analysis. The goal was to select a subset of sensors that explore the 18-dimensional (6 sensors × 3 channels) space generated by all sensors (S1–S6) and simplify that array while maintaining discriminatory capacity. Fig. 13 illustrates the size optimization of the six-member sensor array (Array A). After the optimization process, the PCA score plot still shows clustering with no evident overlap between the different analytes (Fig. 13C). It is also important that this PCA obtained from the 2-sensor array (Array C: S2 and S4) requires 2 dimensions (PCs) out of 6 to describe 94% of the discriminatory range (~40% of all PCs), demonstrating that the reduction of the number of the sensor elements in the array has not significantly affected the discriminatory performance of the array for this data set containing 11 analytes (10 cations and 1 water pH = 5). As expected, the decrease in the magnitude of the eigenvalues (PCs) reveals that the response space generated by the array shrinks due to reducing the number of sensing elements. The decrease in the size of the response space could, however, be contraproductive, *e.g.* in quantitative analysis using reduced array version, because more information is likely to

This journal is © The Royal Society of Chemistry 2010

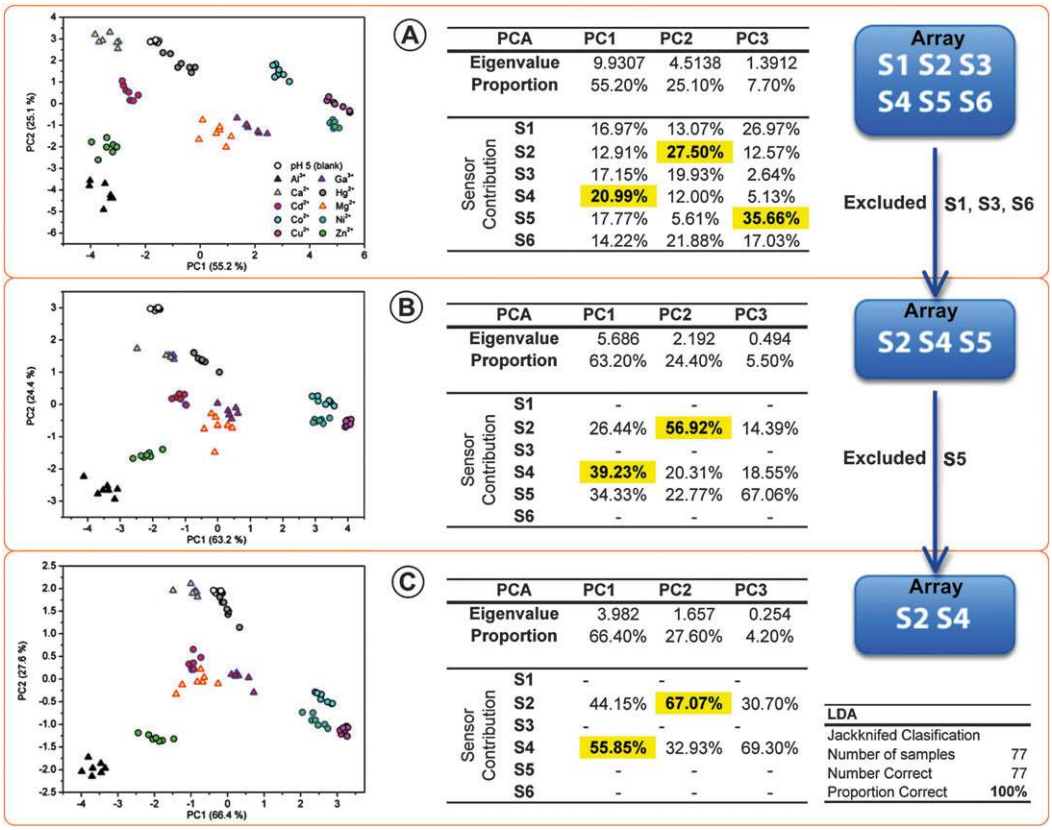

Fig. 13 Schematic representation of the rational process for reduction of the number of sensor elements in an array. From top to bottom: (A) PCA for the complete set of sensors (S1–S6) shows that the main contributors for the dispersion are S4, S2, and S5 on the PCs with statistical significance. (B) Sensors S1, S3, and S5 were excluded from the data set and analyzed again with PCA. PCA shows that the main contributors were S2 and S4. (C) S3 was excluded from the data and PCA was carried out using the remaining data set. Qualitative inspection of the PCA score plot for the final set of two sensors (S2 and S4) shows clustering of the data without any evident overlap between different samples. Cross-validated LDA shows 100% accurate classification for all three arrays (A,B,C). (This figure was reproduced with permission from the American Chemical Society, see ref. 104.)

be needed to avoid overlaps between similar analytes at different concentration. This, again, may be obviated by performing the analysis for each concentration to ensure the satisfactory discrimination is maintained at every concentration.

Hard modeling *vs.* soft modeling

Modeling of sensor arrays is usually based on linear models that are used for calibration and optimization, and utilizes response-surface methods. Using multivariate approaches, the linear models, *i.e.* models linear in the parameters, can be applied to most relationships in supramolecular analytical chemistry including sensor arrays. In addition, when the variables are appropriately transformed (*e.g.* quadratic, cubic, exponential relationship) these models can also be utilized to model nonlinear dependence.

The simplest and well known method is a technique based on ordinary least-squares regression (OLS). In order to introduce soft-modeling techniques, *e.g.* principal component regression (PCR), partial least-squares method (PLS), the OLS formalism will be used for better explanation while focusing on application in array sensors.

Considering optical sensors based on absorption or luminescence, the sensors yield the response as absorbance (A) (Bouger-Lambert-Beer law, (11a),

$$A = \varepsilon \times c \times l \tag{11a}$$

where ε is the molar absorption coefficient, c is analyte concentration, and l the cell path-length, or luminescence intensity (I_f), (11b),

$$I_f = \ln(10) \times \Phi \times I_0 \times \varepsilon \times c \times l \tag{11b}$$

where Φ is the fluorescence quantum yield, and I_0 the intensity of the excitation radiation. In general sense, the optical sensor response, S, is a function of concentration described as

$$S = k \times c \tag{12a}$$

or after rearrangement

$$\log S = \log k + \log c \tag{12b}$$

and the slope of calibration curve ($\boldsymbol{k}$ or $\boldsymbol{\log k}$) is described as the sensor sensitivity.

For the mixture of several optical sensors operating at several wavelengths, or a cross-reactive sensor array, one can write a matrix notation:

$$\boldsymbol{S} = \boldsymbol{C}\,\boldsymbol{K} \tag{13}$$

where $\boldsymbol{S}$ is the $\boldsymbol{N} \times \boldsymbol{K}$ matrix of sensor responses, $\boldsymbol{C}$ is the $\boldsymbol{N} \times \boldsymbol{M}$ matrix of concentrations, and $\boldsymbol{K}$ is the $\boldsymbol{M} \times \boldsymbol{K}$ matrix of regression parameters, $\boldsymbol{N}$ is the number of observations, $\boldsymbol{K}$ is the number of features (see Fig. 8) and $\boldsymbol{M}$ is the number of analytes. Thus in sensor array calibration, this equation can be applied to a model matrix of responses $\boldsymbol{S}$ recorded for features (see Fig. 8) of the respective concentration matrix $\boldsymbol{C}$. In the following text it is assumed that analytical-signal data are centered, *i.e.* there is no intercept calculated for calibration curve (12a). Residuals can be calculated as differences between measured and modeled data, *i.e.* $\boldsymbol{S}$–$\boldsymbol{CK}$.

There are two methods usually applied to this type of problem. The first is the *direct-calibration method* based on OLS, which is used if the K matrix is known assuming that there is no interaction between the different analytes in the sample as well as that no unknown matrix constituents interfere during the analysis. In practice, these prerequisites are not allways fulfilled. Then, more sophisticated calibration procedures have to be applied and multivariate calibration estimating the calibration parameters from calibration mixtures must be used. These *indirect-calibration methods* have some advantages:[3,94]

- In calibration step, the interaction between analytes or analyte(s) and the sample matrix can be taken into account, *i.e.* the validity of eqn (13) is not required.
- Modeling of a background analytical signal considered as a principal component is feasible.
- Systems with highly correlated analytical signal(s) can be also used for multicomponent analysis.

The indirect-calibration procedure consists of two steps: *Calibration*—the matrix $\boldsymbol{S}$ is recorded for calibration set of pure and mixture samples of known concentrations (matrix $\boldsymbol{C}$). Since the matrix of the independent variables, $\boldsymbol{C}$, is not square for so-called over-determined system, the regression parameters $\boldsymbol{K}$ have to be estimated by the generalized OLS inverse method as

$$\boldsymbol{K} = (\boldsymbol{C}^{t}\boldsymbol{C})^{-1}\boldsymbol{C}^{t}\boldsymbol{S} \tag{14}$$

2. *Analysis* of unknown samples is based on the $\boldsymbol{S}_{\text{sample}}$ matrix measurement of sample having an unknown concentration:

$$\boldsymbol{c}_{\text{sample}} = \boldsymbol{S}_{\text{sample}}\,\boldsymbol{K}^{t}(\boldsymbol{K}\boldsymbol{K}^{t})^{-1} \tag{15}$$

Great advantage of this procedure is that the $\boldsymbol{K}$ matrix represents values of optical sensor sensitivities in the presence of all analytes and/or matrix interferents. These values can be compared with the sensitivities obtained from calibration curves in the presence of a single analyte and the selectivity of sensor array can be evaluated comparing both matrices. In hard-modeling approach, all analytes present in the sample should be explicitly known and included in the calibration step. Soft-modeling procedures (*e.g.* PCR, PLS), on the other hand, do not require the knowledge of specific contributions of unknown analytes involved in calibration. In both eqn (14) and (15), a matrix inversion is performed to calculate the matrix of unknown coefficients of the parameters K (from eqn (14)) and a concentration of the unknown sample (eqn (15)). In the case of ill-conditioned (less selective) systems the matrix inversions may be problematic. In such ill-conditioned systems when sensor response "spectrum-like" profiles are very similar, the Gaussian or Gauss-Jordan elimination can be used. Also, more efficient LU-decomposition, Housholder reduction or singular value decomposition (SVD) methods[3,94] are usually used. The currently most frequently used soft-modeling methods (PCR, PLS) are based on the inverse calibration model. The concentration is calculated using regression of sensor response "spectrum-like" profiles

$$\boldsymbol{C} = \boldsymbol{S}\,\boldsymbol{K} \tag{15}$$

These matrices have the same meaning as in eqn (13).

Principal component regression (PCR)

PCR is based on the idea that the principal components obtained by PCA are latent variables which can enter into a (multiple) regression instead of the manifest variables.[30] Linear (in case of one) or multiple linear (in other cases) regressions can be performed using these principal components as the independent variables and the concentrations as dependent variables. Application of techniques for multiple linear regression (MLR) depends on decision regarding how many principal components are included in calibration equation. While all PCs can be utilized in the calculation of regression coefficients, usually only several PCs are employed for calculation.

In an analogy to PCA, one can consider matrix of responses $\boldsymbol{S}$ recorded for sensor features (see Fig. 8) for the respective concentration matrix, $\boldsymbol{C}$. (see Fig. 8). Applying SVD to $\boldsymbol{S}$ data matrix (see also PCA—eqn (7)),

$$\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{t} \tag{16}$$

which is decomposed into two orthonormal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ corresponding to score and loading matrices joined by a diagonal matrix $\boldsymbol{\Sigma}$ of singular values.[3,94] The matrix of the regression coefficients $\boldsymbol{K}$ is performed column-wise as

$$\boldsymbol{K} = \boldsymbol{S}^{+}\,\boldsymbol{C} \tag{17}$$

where the pseudo-inverse matrix $\boldsymbol{S}^{+}$ (Moore-Penrose matrix) is defined as

$$\boldsymbol{S}^{+} = \boldsymbol{V}(\text{diag}(1/\sigma_{\text{ii}})\boldsymbol{U}^{t}) \tag{18}$$

In comparison to MLR, PCR has two main advantages. First, the number of variables is reduced to only a few (feature reduction). Second, the quality of MLR fit is influenced by a degree of correlation between variables and this effect can be eliminated in PCR due to orthogonality of PCs. Other PCs related to additional unknown analyte(s) or background components can be additionally modeled if their concentrations vary within different calibration samples. Sometimes the PCs are neglected as a consequence of their small contributions but they could be important for prediction of a certain concentration. However, since the decomposition of $\boldsymbol{S}$ data matrix sometimes does not match the relationship

This journal is © The Royal Society of Chemistry 2010

(outlined in eqn (13) or (15)). In such a case, the PCR may not give satisfactory prediction of concentration.

Like PCA, also PCR is a linear technique, and therefore it is most successful under assumption of a linear response of the sensors used, for example, in concentration prediction. It should be noted that similarly to LDA, also PCR can give sometimes overly optimistic results, and therefore it is strongly recommended to apply a leave-*n*-out cross-validation procedure or to use external prediction set samples.[95]

Partial least squares (PLS)

PLS regression is a recent technique generalizing and combining features of PCA and MLR. It is also a viable alternative to PCR and as a consequence of application of similar mathematical procedures PLS gives similar information obtained by PCA and PCR. Comparing PLS and PCR, PLS uses more information than PCR does. The inverse calibration model

$$\boldsymbol{C} = \boldsymbol{S}\,\boldsymbol{K} \tag{19}$$

is based on a bilinear model with respect to the objects and the variables of the $\boldsymbol{C}$ or $\boldsymbol{S}$ matrix. Both the $\boldsymbol{C}$ and $\boldsymbol{S}$-matrices are decomposed into smaller matrices according to the following scheme:[3,70,71]

$$\boldsymbol{S} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{t}} + \boldsymbol{E} \tag{20}$$

$$\boldsymbol{C} = \boldsymbol{T}\boldsymbol{Q}^{\mathrm{t}} + \boldsymbol{F} \tag{21}$$

where $\boldsymbol{T}$ is the score matrix containing orthogonal rows, $\boldsymbol{P}$ and $\boldsymbol{Q}$ are the loadings of the $\boldsymbol{S}$ and $\boldsymbol{C}$ matrices, $\boldsymbol{E}$ and $\boldsymbol{F}$ are the error matrices of the $\boldsymbol{S}$ and $\boldsymbol{C}$ matrices. The calculation of the sensitivity $\boldsymbol{K}$ matrix is done using $\boldsymbol{P}$, $\boldsymbol{Q}$ and $\boldsymbol{\Sigma}$ matrices:

$$\boldsymbol{K} = \boldsymbol{\Sigma}(\boldsymbol{P}^{t}\boldsymbol{\Sigma})^{-1}\boldsymbol{Q}^{t} \tag{22}$$

PLS scores and PLS loadings are obtained similarly to PCA procedure and they can be projected in the form of 2D or 3D graphs. Thus PLS extracts more information from experimental data than PCR. In contradistinction to $\boldsymbol{U}$ and $\boldsymbol{V}$ matrices in PCR that are orthogonal, the loadings represented by the $\boldsymbol{P}$ and $\boldsymbol{Q}$ matrices in PLS procedure do not to be orthogonal.[3,94,95] Concentration estimation of unknown samples can be calculated in a way similar to the description in the section on OLS and PCR using Moore-Penrose matrix (see eqn (18)).[3,94,95] The main advantage of the PLS method is that the decomposition of both matrices $\boldsymbol{C}$ and $\boldsymbol{S}$ is interrelated and thus PLS is considered as the most robust calibration algorithm.

The most utilized PLS algorithm is the Nonlinear Iterative Partial Least Squares (NIPALS); details and algorithm descriptions can be found in literature.[94] Comparison of PCR and PLS procedure suggests that both methods are very similar in their outcome.[121] However, PLS appears to have optimal prediction ability with fewer factors than PCR, and therefore PLS appears to be favored over PCR.

It should be noted that PLS, similarly to PCA and PCR, is a linear method, and therefore some precautions should be taken into account. The external prediction set should be used for test of model accuracy in quantitative prediction. Nonlinear variant of PLS regression method was also described.[95,122]

Hierarchical clustering analysis (HCA)[123]

The HCA is an unsupervised method of multivariate analysis, which seeks classification of the observation by measuring the interpoint distances between all samples in the *n*-dimensional space resulting from *n*-numbers of studied features. The observations are aggregated (clustered) stepwise following the similarities (or distances) in their features.

There are several ways to calculate the distance and most of them are included in commercial versions of statistical software. In most cases, chemists use simple HCA and the Euclidean distance (d_{mn}) between the sample m and n (or the square form) with k features as follows:

$$d_{mn}^2 = \sum_{k=1}^{K} (x_{mk} - x_{nk})^2 \tag{23}$$

Also, there are several methods available to define the linkage between the clusters.[93,123] A number of studies use Ward's (minimum variance) method,[124] which takes into consideration the minimum amount of the variance between the samples and analytes to define a cluster. For Ward's method, each first sample is considered its own cluster, therefore, the variance is null. Then, each following step considers a pair of objects that can be grouped while keeping the amount of variance as small as possible. This process is repeated until a supercluster is formed.

The distance (D_{MN}) between two clusters (M and N) is defined by

$$D_{MN} = \frac{\left\|\bar{x}_M - \bar{x}_N\right\|^2}{\frac{1}{n_M} + \frac{1}{n_N}} \tag{24}$$

where n is the number of members in a cluster.

The clusters are then combined in such a way that the minimal increment in the within-group-distance is achieved by making the distance (D_{ji}) between a new cluster j and the observation i, such that the distance from j and i to the clusters M and N results:

$$Dji = \frac{n_M + n_i}{n + n_i} D_{Mi} + \frac{n_N + n_i}{n + n_i} D_{Ni} - \frac{n_i}{n + n_i} D_{MN} \tag{25}$$

Ward's method usually tends to generate well-structured dendrograms (a dendrogram is a mono-dimensional representation that correlates the distances in the *n*-dimensional space between a series of observations). The downside is that it also tends to join different clusters with a small number of observations, and it is strongly biased toward producing clusters with roughly the same number of observations,[125] which is not a problem in most applications if one constantly records the same number of repetitions per analyte.

In the context of chemical sensor arrays, it is important to realize that HCA is very sensitive as it utilizes the whole dimensionality of the data to represent the patterns. The HCA produces dendrograms displaying the quantitative differences (or similarities) between the individual

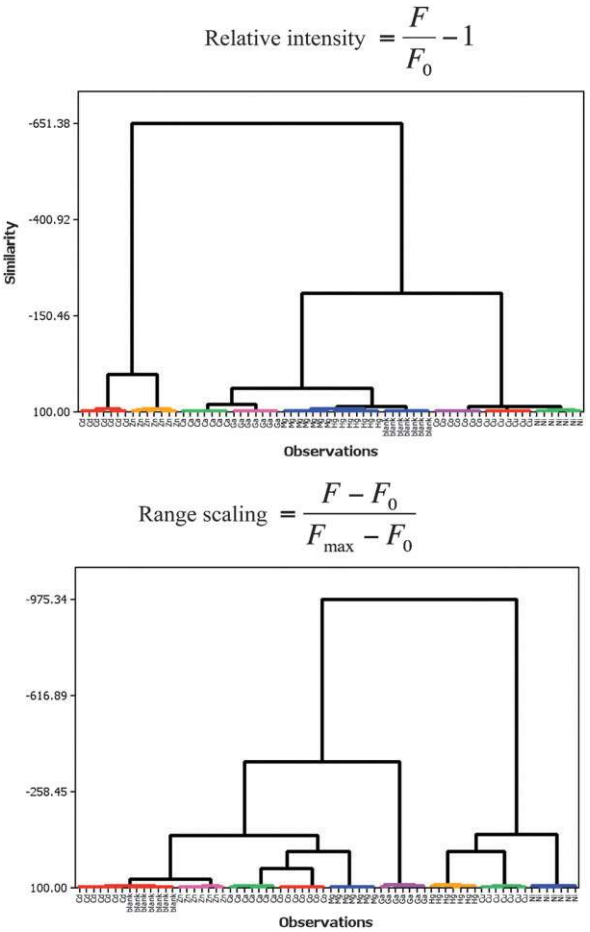

Fig. 14 Impact of the preprocessing of a data set on the HCA (displaying square Euclidean distance with Ward linkages). The data set is composed of the responses generated by the sensor array described in the literature[103] and shows response to 9 different cations (250 μM) at pH 5. Top: Data preprocessed using relative change in the fluorescence intensity. Bottom: Data preprocessed using range scaling algorithm.

observations in a mono-dimensional fashion. This is a valuable information because it makes for a straightforward correlation of chemical features with responses in the *n*-dimensional space. However, in the case when the data set is noisy or does not present a clear structure, the HCA often produces a poor clustering of similar observations.

As in the case of the PCA, the preprocessing of the data is likely to impact the final outcome of the HCA. Here, using the same data set and preprocessing techniques as in the previous example for the PCA (Fig. 9), it is shown how the two preprocessing procedures yield dendrograms with different structures (Fig. 14). Similarly to the PCA, this data set shows better clustering of the data when the range scaling is applied (Fig. 14, bottom).

Current literature describes a number of instances where the optical array sensors were evaluated using HCA. These include samples and analyses of beverages[116] sweeteners[115] saccharides[109,126] and oligosaccharides,[127] amines,[128] and other analytes. The following example (Fig. 15) from the work of Anzenbacher and co-workers shows a colorimetric sensor array for anion detection capable of differentiating between more than 20 analytes including multianalyte samples.[102]

The high-resolution array consisted of eight colorimetric sensors utilizing octamethylcalix[4]pyrrole receptor with selectivity bias toward fluoride and pyrophosphate, but still displaying a cross-reactive response to other anionic analytes. The same array was used for the detection of anions in water and to differentiate toothpaste brands with various content of fluoride and other anions. The data recorded from the array consist of RGB values of each of the 8 sensors to generate a total of 24 variables (8 sensors × 3 colors). HCA was used to show the clustering of the responses of the sensors to 20 different analytes, 10 anions in water as well as identification of toothpastes based on their anion content (Fig. 15).

Supervised pattern recognition and linear discriminant analysis (LDA)

LDA is a classical statistical approach for supervised dimensionality reduction. Indeed, LDA is used for both dimensionality reduction and also for classification (sometimes called supervised pattern recognition). Usually, the first step in supervised pattern recognition is developing a mathematical model describing relationship between trials relating to a series of observations and their known groups. These are called training sets. Once the mathematical model (discriminant function) is developed, it is usually tested as to how well the model predicts the groups. Toward this end, a test set of samples (trials) is used that had been left out during the original calculation that served to develop the model. This process is called validation. Alternatively, a cross-validation approach may be used. Here, only a single training set is used, one sample (trial) is removed at the time and the model is recalculated using the remaining samples. Then the previously removed sample is used to test the model (to predict into which class it belongs). The most common approach is the Leave-One-Out (LOO) (or jackknife) cross-validation where one sample is left out one at a time. The procedure is repeated until all samples have been left out and classified. Alternative to LOO, *v*-folds cross-validation is used in case of large data sets. *v*-Folds takes out a *v*-part of the data set and uses the rest to calculate the model. The process repeats *v*-times and the classification accuracies are averaged.

The majority of statistical treatments seek classification and employ various classification algorithms and procedures such as discriminant analysis and canonical variate analysis. The LDA is also a special case of discriminant analyses where the discriminant rule is based on linear combinations of features (*e.g.*, sensors responses) that best separate two or more analytes. Using the defined group classes, LDA aims to maximize the ratio of the between-the-class distance to the within-the-class distance, thus maximizing the class discrimination. As in PCA, the linear combinations of the features are found by solving an eigenvalue problem.[93,94] As described in the literature,[94] the weights of the linear discriminant functions are determined from the eigenvector of the matrix D

$$D = G^{-1}Hw = \lambda w \quad (27)$$

where λ is the eigenvalue.

This journal is © The Royal Society of Chemistry 2010

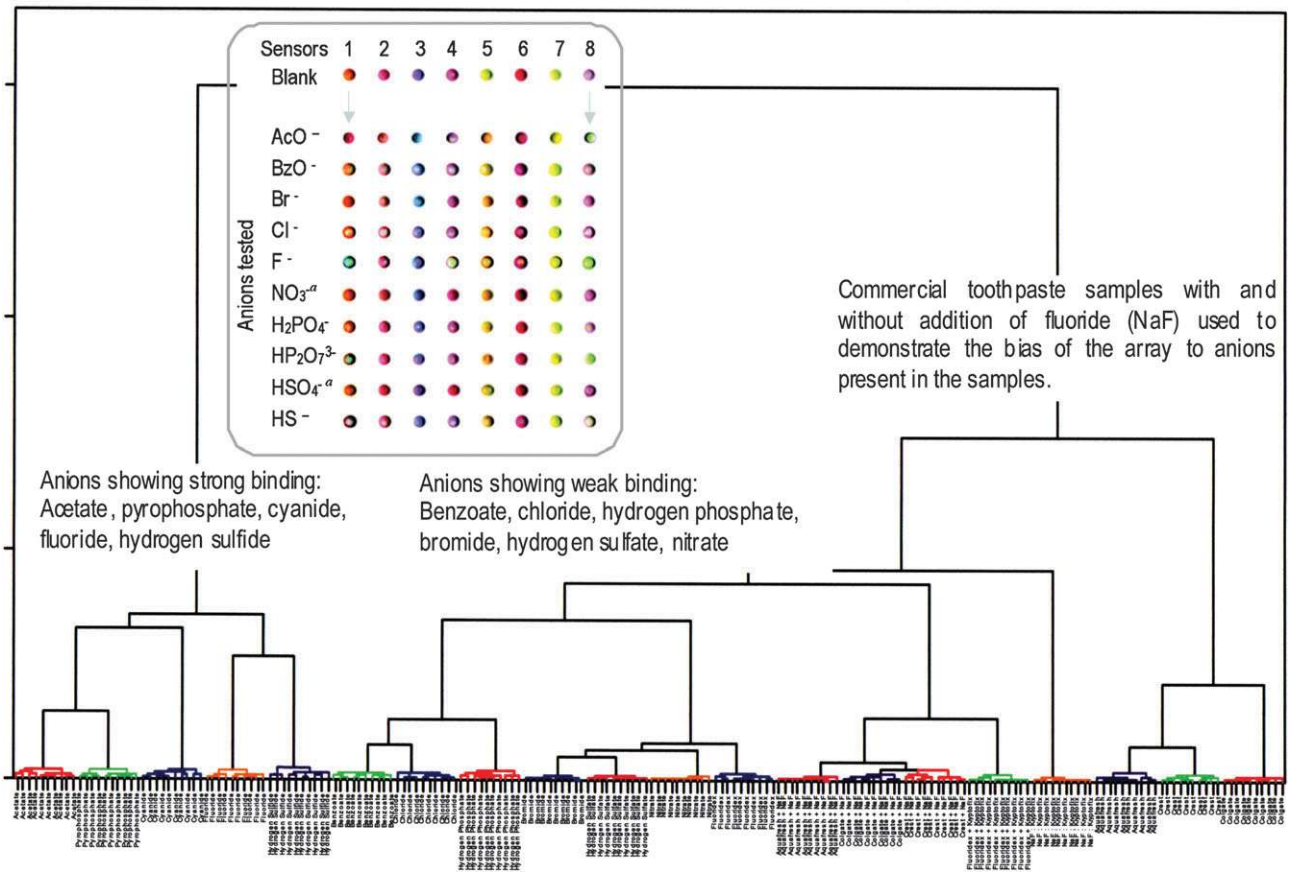

Fig. 15 HCA dendrogram with Ward linkage showing Euclidean distance between 200 samples (11 anions, 4 toothpaste brands and control experiments, 10 trials each analyte). More details are provided in the literature.[102] Inset shows the section of the 8-sensor element array used to recognize 10 different anions.

The matrix G is found from the covariance matrix C of the different groups g following:

$$G = (n-g)C = (n-g)\frac{1}{n-g}\sum_{j=1}^{g}(n_j-1)\boldsymbol{C}_j \qquad (28)$$

and

$$\boldsymbol{C}_j = \frac{1}{n_j-1}\sum_{l\in g_j}(x_{li}-\bar{x}_{ji})(x_{lk}-\bar{x}_{jk}) \qquad (29)$$

for n equal to the total number of observations, n_j equal to number of observation in the group j and l is one observation of the jth group g_j.

The matrix H explains the distribution of the group means g_j over the total average $\bar{x}$,

$$H = \sum_{j=1}^{g} n_j(\bar{x}_j-\bar{x})(\bar{x}_j-\bar{x})^T \qquad (30)$$

$$\bar{x} = \frac{\sum_{j=1}^{n} n_j\bar{x}_j}{n} \qquad (31)$$

Solving the eigenvalue problem $D = G^{-1}Hw = \lambda w$ will derive a list of eigenvalues (λ) and eigenvectors (w). The eigenvector w_1 associated with the greatest eigenvalue λ_1 provides the first discriminant function s_1:

$$s_1 = w_{11}x_1 + w_{12}x_2 + \cdots + w_{1p}x_p \qquad (32)$$

Utilizing the rest of the x-data the following eigenvalue λ_2 is calculated along with eigenvector w_2, to obtain the second discriminant function, s_2:

$$s_2 = w_{21}x_1 + w_{22}x_1 + \cdots + w_{2p}x_p \qquad (33)$$

The calculation of the discriminant functions continues until all are determined. In order to classify any observation, the vector response (in case of sensor arrays) is evaluated in the discriminant functions in order to transform the vector of the raw data in the coordinates within the discriminant space. The observation is then assigned to the group to which it has the minimal Euclidean distance:

$$\min_j ||w^T(x_u-\bar{x}_j)|| \qquad (34)$$

where $j = 1,2,\ldots, g$.

The results of the classification could be represented in a confusion matrix, which is a matrix that contains the numbers of correctly classified objects in each class on the main diagonal and the misclassified objects in the off-diagonal.

The confusion matrix often overestimates the accuracy of the classification due to the bias imposed when a sample

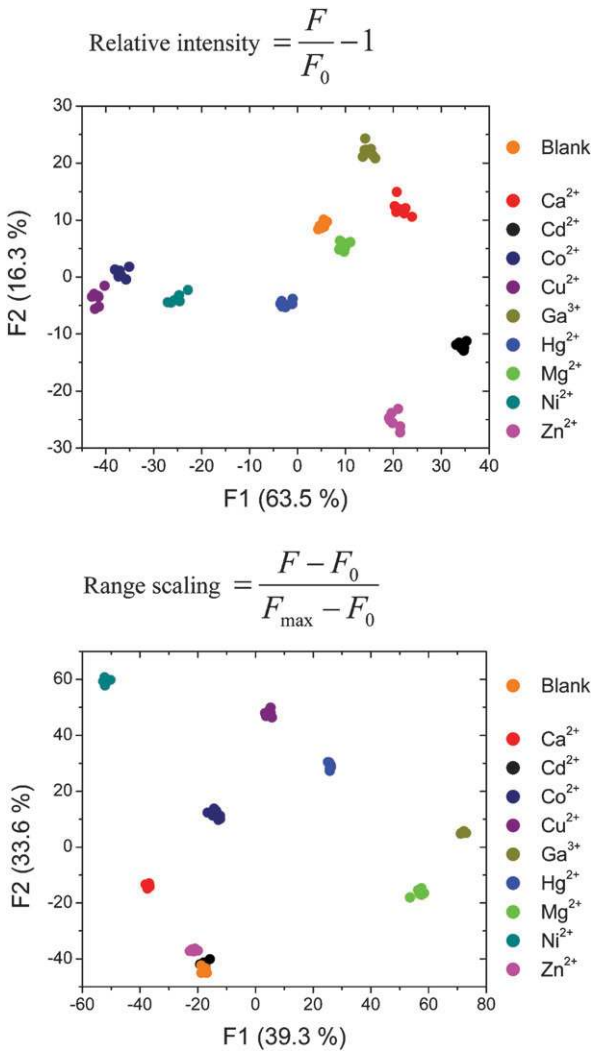

Fig. 16 Impact of the preprocessing of a data set on the LDA. Once again, the data set is composed of the responses generated by the sensor array described in the literature[103] and describe array sensor response to 9 different cations. Top: Data preprocessed using a relative change in the fluorescence intensity. Bottom: Data preprocessed using range scaling algorithm.

classification is attempted while using a training set that contains the same observation. The cross-validation (leave-*n*-out) routine is used to test the predictability of the sensor array by leaving one or more (*n*) observations out of the set at a time, and it uses the rest of the data as a training set to generate the linear discriminant function, which is then used to place the excluded observation (data point) within the correct cluster. This is performed for each observation, and the overall ability to classify the observations describes the quality and predictability of the array. As in the case of the confusion matrix, the cross-validated results can be represented in a jackknifed matrix.

The implementation of LDA is conveniently used for determination of classes by using the linear discriminant function, and it also provides a graphical output by plotting discriminant scores (w_i) against the canonical roots (or factors). These plots provide a graphical representation of how LDA is clustering similar patterns, and it attests to the degree of discrimination of the data, *i.e.*, how good the resolution of the array is for a given group of samples.

As in the case of the two previous analyses the preprocessing of the data is likely to have an impact on the final outcome of an LDA. Once more, the same data set and preprocessing techniques were used to generate two discriminant score plots from the same original response data set as in the previous examples for PCA and HCA. These showed how the two preprocessing data procedures yield score plots with different clustering (Fig. 16). Generally speaking, the LDA shows better clustering of the data when compared to the PCA. Interestingly, in the LDA this data set shows better clustering of the data when the relative intensity is applied (Fig. 16, top).

LDA was successfully used in a number of statistical analyses aimed at classification of analytes.

It was used in a variety of settings and media including solutions, vapors or solid–liquid interfaces such as in the molecularly imprinted polymers.[129] The analytes classified range from small ions and molecules, saccharides,[109,130] amines,[131,132] explosives,[133,134] peptides to evaluate peptide phosphorylation,[135] and many others. In the arena of bio-recognition, LDA was used with enormous success. For example, the surface-modified metal nanoparticles allowed for the analysis and classification of proteins,[136,137] biological fluids,[138] cells including cancerous cells,[139] and bacteria[140] attracted particular attention.

An interesting example of a successful array and LDA analysis was recently published by Lavigne *et al.*,[131,132] who used a polymer of 3-(thiophen-3-yl)propanoic acid for classification of various aliphatic as well as aromatic amines, diamnines and polyamines (Fig. 17). The highly cross-reactive conjugated polythiophene bearing negatively charged carboxylate residues forms ammonium salts with the analytes. The resulting aggregation of the polythiophene-amine results in a change in color, which is clearly observable by UV-vis spectroscopy and is easily performed in multi-well plates. The single polythiophene (Fig. 17A) was shown to generate a multidimensional response useful for classification of 22 structurally similar and biologically relevant amines with 97% accuracy. Similar multidimensional response from the cross-responsive poly(thiophene) has been analyzed using a wavelength ratiometric method to quantify the amount of biogenic amine present in a fish meat matrix, thereby evaluating the quality of the food. This method allowed for identification of amines such as putrescine, cadaverine and histamine that are associated with food decomposition, and meat in particular.

In another example, LDA was successfully used to analyze the responses of a 9-member sensor array in the identification of mineral and purified (Aquafina) water samples based on their cation content (mostly Ca^{2+}, Mg^{2+}).[103] Anzenbacher and co-workers studied the responses for nine commercial brands of mineral water along with two controls/blanks. Fig. 18 lists the calcium and magnesium ion contents for all of the mineral water brands tested. From the list (Fig. 18, top) it is clear that all nine brands contain different kinds and concentrations of cations and also in different proportions. LDA coupled with leave-one-out cross-validation routine

This journal is © The Royal Society of Chemistry 2010

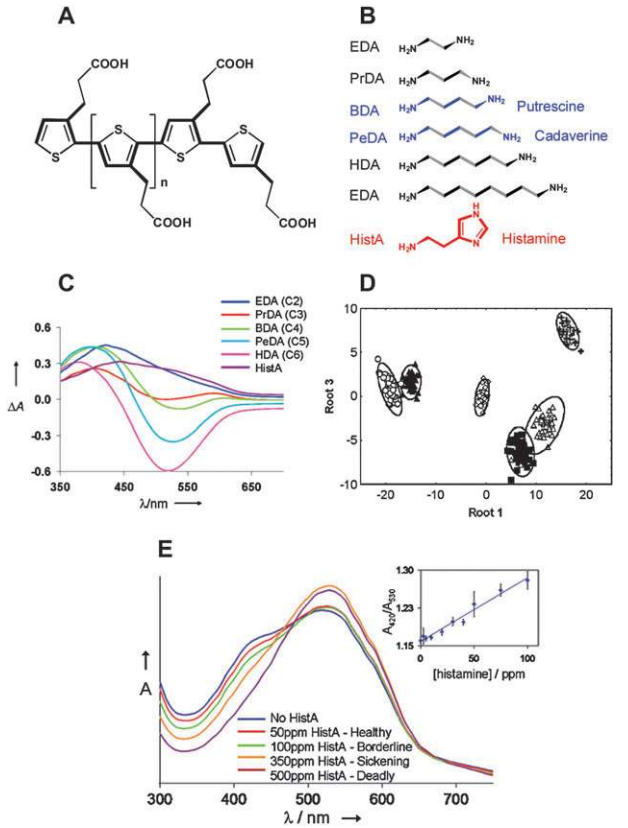

Fig. 17 Schematic representation of the cross-reactive thiophene polymer (A) and examples of amines (B) shown to display analyte-induced aggregation. Unique optical signatures (C) represented by the changes in the absorption spectra of the polymer chemosensor (1 mM) upon addition of different diamines (1 mM each) in aqueous HEPES buffer (40 mM, pH 7.4). Panel D: Two-dimensional LDA plot of chemosensor response to EDA (■), PrDA (△), BDA (▲), PeDA (◇), HDA (○) and HistA (+). Panel E: Response of the polymer chemosensor Polymer **1** response to histamine in a tuna fish matrix. Inset: Response factor (A_{420}/A_{530}) showing linearity in the polymer response in the relevant range to detect spoilage. (Fig. 17C–E were reproduced with permission from the American Chemical Society, see ref. 132.)

Metal Ion Content (mg/L)					
Mineral Water	**Ca^{2+}**	**Mg^{2+}**	**Na^{+}**	**K^{+}**	**pH**
Aquafina	~0	~0	~0	~0	7.1
Biovive	42	3.8	2.5	0.6	7.5
Bonaqua	62	41	118	10.8	6.1
Evian	78	24	5	1	7.3
Fiji	17	13	17	1	7.5
Gerolsteiner	348	108	118	10.8	6.6
Magnesia	38	234	5	1	7.0
Naleczowianka	114	23	13	5	7.5
Perrier	149	7	12	1	5.1

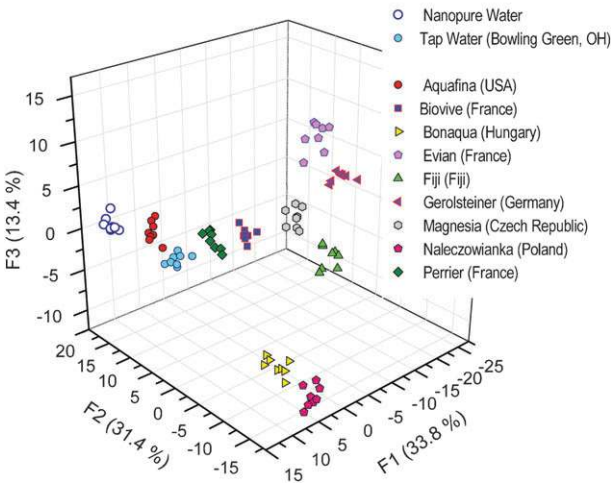

Fig. 18 Metal ion content for different brands of mineral and purified water samples. Right: LDA canonical score plots corresponding to the response of the 9-member array to 9 different water brands. (This figure was reproduced with permission from the American Chemical Society, see ref. 103.)

showed 100% correct classification for all 88 trials (Fig. 18, bottom). Interestingly, Aquafina® brand, due to its low electrolyte content, presents a very weak response, owing to the fact that it is closest to the nanopure water by cation content; as shown on the LDA canonical score plot. Aquafina® is actually not commercialized as a "mineral water," but as "pure water."

Artificial neural networks

The number of papers on applications of Artificial neural networks (ANN) have grown significantly since 1990s. Although the original excitement about ANNs has somewhat waned and a more nuanced view of these methods has developed, there are still issues, which are often overlooked. ANNs are often confused with the model of natural brains. It should be noted that artificial neural networks have only little similarity to natural biologic neural networks.

The users of artificial neural networks often regard ANNs as a type of a black magic box which can be used to enter data and get a solution back. Although this view is potentially dangerous, there is also a grain of truth in it. Artificial neural networks (ANN) are adaptive models that can establish almost any relationship between data. They can be regarded as black boxes to build mappings between a set of input and output vectors. ANNs are particularly well suited for solving problems where traditional models fail, especially for modeling complex phenomena which show a non-linear relationship (Fig. 19). One could therefore regard an ANN as an abstract machine which creates a non-linear mapping between an n-dimensional input data space and a *p*-dimension output space. **n** is usually much larger than **p**; with **p** often

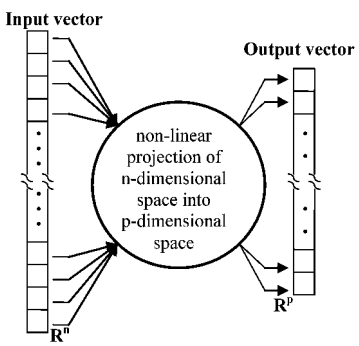

Fig. 19 General outline of an artificial neural network (ANN).

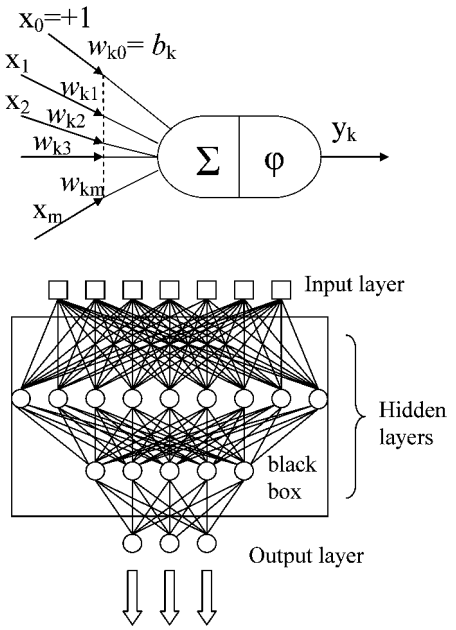

Fig. 20 Considering the operation of a single neuron k. It receives the input signals x_m from other neurons 1-m, aggregates these signals (Σ) by using the weights of the synapses (w_{km}) and after suitable transformation (φ) passes the result as the output signal y_k.

being in the range of 1 to 3 (since the human interpreter is restricted to a maximum dimensionality of 3).

This non-linear mapping is set up during the learning process of a neural network. The "art" of training a neural network is to control the training in such a way that the resulting mapping represents the underlying relationship within the data, avoiding adjustment for noise or errors in the data.

In order to select the appropriate neural network model, it is necessary to realize whether one is dealing with a supervised task, or unsupervised task. To train the system to respond with a certain predefined output, the task is supervised. If the network should self-organize, the target is not predetermined and the task is called an unsupervised task (see above).

Similarly, it is important to realize whether a task is a classification task or function approximation task where each input is associated with a certain (analog) value. Supervised training methods use training sets to set up a relationship between input and output of the ANN model. Supervised methods are therefore mostly used for function approximation and classification, while unsupervised methods are most suitable for clustering tasks. Unsupervised methods try to find the structure in the data on their own.

The central paradigm of neural networks is based on local computing. This means that neural networks gain their power from connecting several processing units, which have only very restricted capabilities. A multi-layer ANN consists of units and connections. Each unit requires an activation, and each link between two units has a weight. The units are organized in layers. Three different types of units are distinguished and are divided into input units, hidden units, and output units. Fig. 20 shows a simple type of a feed-forward artificial neural network.

The basic idea of ANN is that the information passed along the path (an analogy of axon in the actual neuron) arrives at the synapse. There, the information is transformed and sent to the next neuron. This signal transfer is simulated in the artificial neuron considering the input signal x_n, modified with the synaptic weight w, to derive the output signal y (Fig. 20), where the output is defined $y = wx$. For a given artificial neuron, the $m + 1$ are inputs with signals x_0 through x_m and weights w_0 through w_m. Usually, the x_0 input is assigned the value $+1$, which makes it a *bias* input with $w_{k0} = b_k$. This leaves only m actual inputs to the neuron: from x_1 to x_m. The output of kth neuron is:

$$y_k = \varphi\left(\sum_{j=0}^{m} w_{kj}x_j\right) \tag{35}$$

and where φ is the transfer function.

Between the input and output layer additional layer(s) called hidden layers, may be present. When considering the neural network as a black box, the hidden units are not visible from the outside (Fig. 20, right). The calculation of the final output values proceeds layer by layer. First, the input signals are applied to the input layer, and each neuron of the input layer calculates its output value.

Next, these values are propagated to the next layer and so forth, until the output layer is reached. Each unit uses the output of the previous units as an input and provides its output as input for the subsequent units (hence the term feed-forward ANN). The coefficients w_{kj} that multiply the inputs x_j are called weights. The process of training a network consists of adjusting the weights to minimize disagreement between the output of the network and desired values for a training set with known correct outputs. Thus, the network "learns" and the w_{kj} are adjusted in the process. Standard ANN used in chemistry usually comprises a hierarchy of identical units.

The unit multiplies each input by the weight of the connection between itself and the unit providing the input. It adds together the weighted inputs and bias term ($x_0 = +1$, thus x_0 $w_0 = b_k$). The bias term can be regarded as the weight of an input fixed to 1.

$$\mathrm{Unit}_k = \mathrm{b}_k + \varphi\left(\sum_{j=1}^{m-1} w_{kj}x_j\right) \tag{36}$$

Typically, the aggregation of the input signals (x_m w_km) is a simple summation. Other aggregation operators are also possible (*e.g.* formation of minimum or maximum over all signals). Before the aggregated signal leaves the unit, a transformation is performed by applying a transfer function (φ). Variety of transformation functions can be used, for example, threshold logic, linear transfer function, sigmoid function, linear threshold function, hyperbolic tangent function, *etc.*

There are several neural network models commonly used by chemists. The simple neural network is the perceptron. Architectures more frequently used by chemists are backpropagating networks and Kohonen networks, which are used for unsupervised learning and classification. Thus, the choice of architecture depends on the analytical objective. For example,

 This journal is © The Royal Society of Chemistry 2010

for parameter estimation backpropagation networks are usually used, while for the clustering analysis the Kohonen networks appear to be preferable. Backpropagation networks are also used for classification and identification applications.

This procedure is well documented in many publications and more details including the introduction into the subject are given in the literature.[141,142] While the artificial neural networks (ANNs) are extensively applied to electrochemical sensor arrays in recent studies,[26,143] there are only a few reports describing utilization in arrays of optical sensors. These are mostly used for multi-component analysis utilizing newly synthesized receptors and used for determination of analyte concentrations[152,153] in solution or by means of sensor arrays[150] or array of fiber-optics sensors.[148,144] Severin *et al.* applied dynamic library of organometallic receptor and three azodyes followed by ANN evaluation of experimental data for determination of concentration of ATP and AMP/pyrophosphate.[145,146]

It appears that there is no description of ANN applied to the classification in optical sensor arrays. To create such an example for this review, we used the experimental data from Palacios and co-workers,[104] which is discussed in detail in the section devoted to PCA and the procedure for the sensor array size optimization. The same data-set was analyzed using back-propagating ANNs. The data set describes array responses as inputs and elements (in nominal or alphanumeric form) as outputs for only one concentration (1 mM, pH = 5) and comprises 7 repetitions. The 7 repetitions are split randomly into three sets of 4, 2 and 1 repetitions:

The first portion of the data (first 4 repetitions) recorded for 10 ions (Al, Ca, Cd, Co, Cu, Ga, Hg, Mg, Ni, Zn) and a blank yields 44 data points and is used as a training set for the next step, which is analysis, namely for search for a minimum of residual function $U = \Sigma\ (\mathrm{Output}_{exp} - \mathrm{Output}_{calc})^2$, applied over all experimental points. The well known optimization procedures such as the conjugate gradients method or Levenberg-Marquardt approach may be applied. Also some other parameters (*e.g.* learning rate, momentum, *etc.*) do have influence on speed of the optimization process and they can be critical in the search for global minimum. The second portion (verification data set) of the data (two out of remaining three repetitions for 10 ions Al, Ca, Cd, Co, Cu, Ga, Hg, Mg, Ni, Zn and a blank), gives 22 data points, which are then used to optimize the ANN architecture, particularly the number of hidden neurons (Table 2) and other parameters (weights, biases) to achieve the best classification. The third and last set, comprised of the last repetition gives 11 data points and serves to test the suitability and accuracy of the prediction.

This is done simply by performing the classification of these data points using the previously trained neural network with the optimized architecture.

Finally, in order to evaluate the effect of preprocessing the experimental data, two data sets for testing were analyzed. The preprocessing method used was a simple average. This way, another "backup" set of 11 data points was created as an average value of the 7 repetitions. The backup data set (composed of the averages of the 7 repetitions) will be used in the case the last remaining repetition (the 11 data points) are biased by an outliers, a systematic error, *etc.* This preprocessed data set reflects the properties of all the repetitions, and therefore is likely to show lower outlier bias than the individual repetition data points.

The results are shown in Table 2. For a full array comprising all 6 sensors (S1–S6), one can obtain good prediction for both data sets (91% and 100% success in prediction). Decreasing the number of sensors in an array to half, the success in the prediction remains unchanged (91% and 100%), which is the same as the classification accuracy described for LDA in the study by Anzenbacher *et al.*[104] Note, however, that the LDA utilized all seven repetitions and a leave-one-out (LOO) accuracy test. Thus, the current ANN analysis actually utilizes only a very small fraction of the data set to achieve very similar classification accuracy. The 3-sensor array is sufficient for practical classification of 10 analytes (metal ions) and distinguishes them from a blank.

In the case of two-member arrays (S2 and S4), the correct prediction is achieved only for the data set preprocessed as an average. Here, the prediction for random data (7th repetition) was less successful. Specifically, Ca and Hg were misclassified as blanks, Cd as Mg, and Mg as Cd. Interestingly, when the preprocessed data points (generated as an average of repetitions, see above) were used, a 100% of correct classification was achieved. Thus, it appears that the original experimental data suffer from an error, for example, due to noisy data, or a presence of a systematic error. However, such an effect can be successfully compensated for by a simple pretreatment such as signal averaging. This illustrates the

Table 2 The results of ANN analysis. The calculations have been carried out with TRAJAN® software v 3.0. For experimental details see the literature[104]

Number of sensors[a]	Neural Network Architecture[b] RMS (training and verification set)	RMS (test data set)[c] Number of samples/Number of correct classifications
6 (S1–S6)	18 : 13 : 11	0.0235 (11/10, *i.e.* 91%)
	(0.014; 0.029)	0.0089 (11/11, *i.e.* 100%)
3 (S2, S4, S6)	9 : 18 : 11	0.0265 (11/10, *i.e.* 91%)
	(0.019; 0.031)	0.0087 (11/11, *i.e.* 100%)
2 (S2, S4)	6 : 16 : 11	0.0708 (11/7, *i.e.* 64%)
	(0.026; 0.045)	0.0115 (11/11, *i.e.* 100%)

[a] Three emission channels were used: blue, green, yellow. [b] ANN inputs : hidden neurons : outputs, Activation function—logistic, PSP function—linear; learning rate 0.8–0.9; momentum 0.3; Training data—44 points, Verification data—22 points. [c] Test data—11 points (Blank, Al, Ca, Cd, Co, Cu, Ga, Hg, Mg, Ni, Zn). 1st randomly selected data set corresponding to the last repetition, 2nd data preprocessed set ("back-up") obtained by averaging of 7 repeated measurements.

importance of careful examination of the experimental data prior to advanced chemometric analysis, but also the overall robustness of the ANN method for data classification.

There are other examples in the literature of using simple ANN to perform the analyses of the data from optical sensor arrays. Walt and co-workers demonstrated a cross-reactive optical sensor array, a true multi-analyte fiber-optic sensor modeled directly on the olfactory system for sensing of vapors.[84,147,148] Also, Jurs and co-workers developed an optical fiber array sensor for organic volatiles that also utilizes ANN.[147,149] Sensing of heavy metal cations in ternary mixtures from absorption spectra of three types of commercially available metallochromic indicators using ANN were performed by Suzuki *et al.*[150] Similar work on recognition of metal cations for the purpose of environmental monitoring was also reported.[151] Additionally, Anslyn *et al.* used multilayer perceptron (MLP) to fingerprint chemical identity, chirality, and concentration of four chiral diamines, 1,2-diphenylethylenediamine, 1,2-diaminocyclo-hexane, 1,2-diaminopropane, bis-(4-methoxyphenyl)-1,2-di-aminoethane using exciton-coupled circular dichroism spectroscopy. Copper(I) 2,2′-bis(diphenylphosphino)-1,1′-dinaphthyl-acetonitrile complex $[Cu(I)(BINAP)(MeCN)_2]PF_6$ in the form of its both enantiomers (*R* and *S*). The copper-BINAP complex displays a broad metal-to-ligand charge transfer band at 350–430 nm with opposite sense in opposite enantiomers. Association of the chemosensor enantiomers with the chiral diamine resulted in an analyte-specific change in the CD spectra of the MLCT region resulting from the formed chemosensor-diamine complex. A similar method was then applied to enantio-selective indicator displacement assays for R-amino acids to achieve high-throughput screening of enantiomeric excess. Complexes of chiral diamines with Cu(II) were incubated with the color indicator chrome azurol, and the dye displacement as a function of enantiomeric excess was analyzed using ANN as well as using calibration plots.[152] Similar methods aimed at simplification of enantioselective indicator displacement assays were also published.[153]

Support vector machines

Along with PCA and LDA, there are several other vectorial models for multivariate clustering and data mining, such as factorial analysis, partial least square (PLS)[95] and support vector machines (SVM),[154–157] among others. In the recent years, SVM has appeared as a very reliable method when dealing with data sets with several distinct classes that cannot be easily separated by linear boundaries.[158,159]

Briefly, SVM is a supervised classification method that seeks to separate classes by mapping the input into an *n*-dimensional vector space using kernel functions. These kernel functions can be linear, polynomial, radial basis function, *etc*. The kernel functions give flexibility in terms of the nature of the data to be classified. For example, in case of a data set where the input does not show a linear correlation with the variables or when the data is not linearly separable the kernel functions can be used to map the input (data points) into a feature space where the classes can be linearly separated (Fig. 21A). Once in the feature space, data points are separated by hyperplanes with

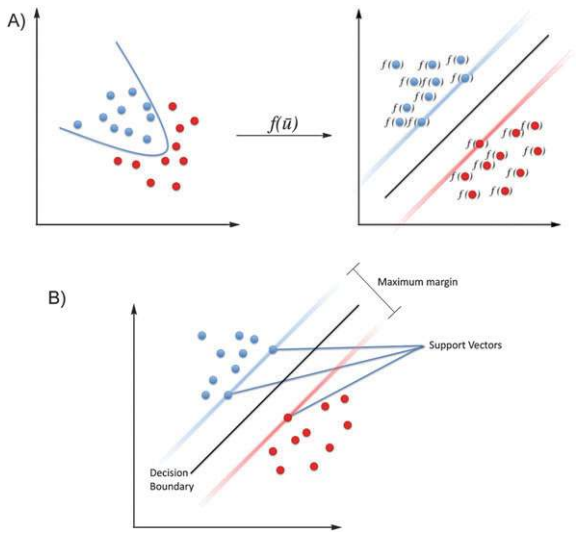

Fig. 21 Representation of the mapping of the data set into a feature space where the data are linearly separable using a kernel $f(\vec{u})$ (A). Support vector machine (B): The hyperplanes comprising support vectors represent the maximum separation margin between two different classes.

maximum soft margin in reference to the training data. The distance between two hyperplanes separating two different classes is called the maximum margin hyperplane and it corresponds to the decision boundary that maximizes the resolution between the two different classes. Thus, SVM looks for data points (support vectors) that satisfied the condition where two hyperplanes display maximum separation between the two classes. Any variation on the support vectors will immediately alter the nature of the hyperplanes defining the margin (Fig. 21B).

Further explanation of SVM goes beyond the scope of this review, but in the context of sensor arrays, SVMs have proved outstanding performance in difficult analytical problems, such as multianalyte samples. From our experience, SVMs usually provide better classification accuracy than linear supervised methods such as LDA. However, they are difficult to implement owing to their intrinsic complexity due to parameterization that most of the time is carried out along with cross-validation routings. SVMs were successfully applied to optical arrays aimed at alkaloids classification using DNA three way junctions,[160] metal ions,[158] organic vapors,[159] and other analytes. Comparing LDA and SVM, as a more sophisticated boundary method, SVM is prone to overfit the normally-distributed data but it generally performs better in case of experimental data with complicated structure.[161]

Conclusions

There is no doubt that the field of sensor arrays will grow and thrive in the near future. There are multiple reasons to justify this prediction. Firstly, the state of the art in analytical and supramolecular chemistry, and synthesis methodology create environment in which array sensors became a mainstream tool of evaluation of numerous processes that transcend the classical analytical application. Secondly, the development of

 This journal is © The Royal Society of Chemistry 2010

micro-instrumentation reached such a state that most high-throughput analysis tools (high-density multi-well readers, imagers, CCD cameras, *etc*.) are widely available and their cost is acceptable for both the industry and academic research groups. This means that more and more researchers are becoming familiar with performing experiments and analyses using array methodologies and will rely on array sensors, methods of analysis, exploration and visualization. Thirdly, the significant motivation exerted by the current economic pressures will, in fact, ensure that efforts toward assay and testing processes in a highly economic fashion will not only espouse the high-throughput hardware (multi-well formats, imaging, *etc*.) but also intellectual tools including applied mathematics and statistics, which will continue to be at the forefront of the R&D efforts of the chemistry and industry.

This climate will create a breeding ground for new types of chemistry professionals. The traditional chemist that used to rely on round-bottom flask reaction testing, column chromatography, HPLC and NMR will be partly replaced by this new chemist trained to utilize small-formate assays including various lab-on-a-chip[162,163] and micro-technologies,[164] to develop tools that allow visualization of the chemical processes *via* imaging as a method of following the processes and multivariate analyses. Even more importantly, the new generation of scientists will use the array tools including the array sensors to monitor relevant processes and analyze the products, production and waste streams and will be more attuned to approach the experiments using careful design and modeling as well as statistical evaluation, data-mining and mathematical predictive methods to minimize the number necessary experiments or elements while maximizing the extracted information.

While some might argue that the combinatorial chemistry[165,166] or nanochemistry[167] did not deliver yet on some of their promises, from today's perspective it is very likely that the sensor arrays and nanotechnology will intersect in the near future. The miniaturization of sensor arrays down to the nanoscale will most likely yield high density arrays to serve reliable sensing processes.[168] Finally, the current state of the art in single molecule analysis suggests that we might not be too far from arrays of single molecules as sensors.[169] Two things are clear: Whilst there are numerous challenges in the field of arrays and array sensors, there are also exciting developments and promising science and technology waiting in the near future.

Notes and references

1 R. W. Cattral, *Chemical Sensors*, Oxford University Press, Oxford, 1997.
2 A. Hulanicki, S. Glab and F. Ingman, *Pure Appl. Chem*., 1991, **63**, 1247.
3 J.-M. Mermet, R. Kellner, M. Otto, M. Valcarcel and M. Widmer, *Analytical Chemistry: A Modern Approach to Analytical Science*, Wiley, New York, 2004.
4 *Microarray Technology and Its Application*, ed. U. R. Muller and D. V. Nicolau, Springer Verlag, Berlin-Heidelberg, 2005.
5 N. Mocellin, Microarray Technology and Cancer Gene Profiling, *Adv. Exp. Med. Biol*., 2007, vol. 593.
6 *Protein Microarrays*, ed. M. Schena, Jones and Bartlett Publishers, Sudbury, MA, 2005.
7 M. Schena, D. Shalon, R. W. Davis and P. O. Brown, *Science*, 1995, **270**, 467.
8 M. Schena, *BioEssays*, 1996, **18**, 427.
9 M. Schena, D. Shalon, R. Heller, A. Chai, P. O. Brown and R. W. Davis, *Proc. Natl. Acad. Sci. U. S. A*., 1996, **93**, 10614.
10 D. Shalon, S. J. Smith and P. O. Brown, *Genome Res*., 1996, **6**, 639.
11 R. A. Heller, M. Schena, A. Chai, T. Bedillion, J. Gilmore, D. E. Woolley and R. W. Davis, *Proc. Natl. Acad. Sci. U. S. A*., 1997, **94**, 2150.
12 R. Forozan, R. Karhu, J. Kononen, A. Kallioniemi and O.-P. Kallioniemi, *Trends Genet*., 1997, **13**, 405.
13 P. T. Spellman, G. Sherlock, M. Q. Zheng, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein and B. Futcher, *Mol. Biol. Cell*., 1998, **9**, 3273.
14 J. Ollila and M. Vihinen, *Biochem. Biophys. Res. Commun*., 1998, **249**, 475.
15 P. A. Outinen, S. K. Sood, P. C. Liaw, K. D. Sarge, N. Maeda, J. Hirsh, J. Ribau, T. J. Podor, J. I. Weitz and R. C. Austin, *Biochem. J*., 1998, **332**, 213.
16 Z. Bryant, L. Subrahmanyan, M. Tworoger, L. LaTray, C.-R. Liu, M.-J. Li, G. Van Den Engh and H. Ruohola-Baker, *Proc. Natl. Acad. Sci. U. S. A*., 1999, **96**, 5559.
17 L. H. Augenlicht, M. Bordonaro, B. G. Heerdt, J. Mariadason and A. Velcich, *Ann. N. Y. Acad. Sci*., 1999, **889**, 20.
18 S. D. Ginsberg, S. E. Hemby, V. M. Lee, J. H. Eberwine and J. Q. Trojanowski, *Ann. Neurol*., 2000, **48**, 77.
19 I. Blumcke, A. J. Becker, S. Normann, V. Hans, B. M. Riederer, S. Krajewski, O. D. Wiestler and G. Reifenberger, *J. Neuropathol. Exp. Neurol*., 2001, **60**, 984.
20 P. J. Ross, M. George, D. Cunningham, F. DiStefano, H. Andreyev, N. Jervoise, P. Workman and P. A. Clarke, *Mol. Cancer Ther*., 2001, **1**, 29.
21 B. Schweitzer and S. F. Kingsmore, *Curr. Opin. Biotechnol*., 2002, **13**, 14.
22 K. R. Love and P. H. Seeberger, *Angew. Chem., Int. Ed*., 2002, **41**, 3583.
23 K. S. Lam and M. Renil, *Curr. Opin. Chem. Biol*., 2002, **6**, 353.
24 R. Z. Wu, S. N. Bailey and D. M. Sabatini, *Trends Cell Biol*., 2002, **12**, 485.
25 M. S. Fejzo and D. J. Slamon, *Am. J. Pathol*., 2001, **159**, 1645.
26 Y. Vlasov, A. Legin, A. Rudnitskya, C. Di Natale and A. D'Amico, *Pure Appl. Chem*., 2005, **77**, 1965.
27 B. E. Collins, A. T. Wright and E. V. Anslyn, *Top. Curr. Chem*., 2007, **277**, 181.
28 D. C. Harris, *Quantitative Chemical Analysis*, W. H. Freeman, New York, 2003.
29 D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte and L. Kaufman, *Chemometrics: A Textbook*, Elsevier, Amsterdam, 1987.
30 *Handbook of Chemometrics and Qualimetrics*, ed. D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, Y. Michotte and L. Kaufman, Elsevier, Amsterdam, 1997.
31 A. Buryak and K. Severin, *Angew. Chem., Int. Ed*., 2005, **44**, 7935.
32 A. Buryak and K. Severin, *J. Am. Chem. Soc*., 2005, **127**, 3700.
33 F. Zaubitzer, A. Buryak and K. Severin, *Chem.–Eur. J*., 2006, **12**, 3928.
34 A. Buryak and K. Severin, *J. Comb. Chem*., 2006, **8**, 540.
35 F. Zaubitzer, T. Riis-Johannessen and K. Severin, *Org. Biomol. Chem*., 2009, **7**, 4598.
36 S. Rochat, J. Gao, X. H. Qian, F. Zaubitzer and K. Severin, *Chem.–Eur. J*., 2010, **16**, 104.
37 S. C. McCleskey, M. J. Griffin, S. E. Schneider, J. T. McDevitt and E. V. Anslyn, *J. Am. Chem. Soc*., 2003, **125**, 1114.
38 S. H. Shabbir, L. A. Joyce, G. M. da Cruz, V. M. Lynch, S. Sorey and E. V. Anslyn, *J. Am. Chem. Soc*., 2009, **131**, 13125.
39 H. S. Hewage and E. V. Anslyn, *J. Am. Chem. Soc*., 2009, **131**, 13099.
40 A. T. Wright, M. J. Griffin, Z. Zhong, S. C. McCleskey, E. V. Anslyn and J. T. McDevitt, *Angew. Chem., Int. Ed*., 2005, **44**, 6375.
41 A. T. Wright, E. V. Anslyn and J. T. McDevitt, *J. Am. Chem. Soc*., 2005, **127**, 17405.
42 S. Nieto, J. M. Dragna and E. V. Anslyn, *Chem.–Eur. J*., 2010, **16**, 227.

43 Y. L. Liu, M. A. Palacios and P. Anzenbacher, Jr., *Chem. Commun*., 2010, **46**, 1860.
44 M. A. Palacios, Z. Wang, V. A. Montes, G. V. Zyryanov and P. Anzenbacher, Jr., *J. Am. Chem. Soc*., 2008, **130**, 10307.
45 M. A. Palacios, R. Nishiyabu, M. Marquez and P. Anzenbacher, Jr., *J. Am. Chem. Soc*., 2007, **129**, 7538.
46 N. A. Rakow, A. Sen, M. C. Janzen, J. B. Ponder and K. S. Suslick, *Angew. Chem., Int. Ed*., 2005, **44**, 4528–4532.
47 M. C. Janzen, J. B. Ponder, D. P. Bailey, C. K. Ingison and K. S. Suslick, *Anal. Chem*., 2006, **78**, 3591.
48 S. H. Lim, Ch. J. Musto, E. Park, W. Zhong and K. S. Suslick, *Org. Lett*., 2008, **10**, 4405.
49 Z. Yao, X. Feng, W. Hong, Ch. Li and G. Shi, *Chem. Commun*., 2009, 4696.
50 T. Mayr, Ch. Igel, G. Liebsch, I. Klimant and O. S. Wolfbeis, *Anal. Chem*., 2003, **75**, 4389.
51 A. Schiller, R. A. Wessling and B. Singaram, *Angew. Chem., Int. Ed*., 2007, **46**, 6457.
52 N. T. Greene, S. L. Morgan and K. D. Shimizu, *Chem. Commun*., 2004, 1172.
53 T. L. Nelson, I. Tran, T. G. Ingallinera, M. S. Maynor and J. J. Lavigne, *Analyst*, 2007, **132**, 1024.
54 A. Bajaj, O. R. Miranda, R. Phillips, I.-B. Kim, D. J. Jerry, U. H. F. Bunz and V. M. Rotello, *J. Am. Chem. Soc*., 2010, **132**, 1018.
55 K. Eckschlager and K. Danzer, *Information Theory in Analytical Chemistry*, Wiley, New York, 1994.
56 S. L. Wiskur and E. V. Anslyn, *J. Am. Chem. Soc*., 2001, **123**, 10109.
57 C. Zhang, D. P. Bailey and K. S. Suslick, *J. Agric. Food Chem*., 2006, **54**, 4925.
58 E. V. Anslyn, *J. Org. Chem*., 2007, **72**, 687.
59 J. J. Lavigne and E. V. Anslyn, *Angew. Chem., Int. Ed*., 1999, **38**, 3666.
60 D. L. Eastwood, O. O. Soyemi, J. Karunamuni, L. Zhang, H. Li and M. L. Myrick, *Proc. SPIE-Int. Soc. Opt. Eng*., 2001, **4199**, 105.
61 D.-S. Lee, J.-K. Jung, J.-W. Lim, J.-S. Huh and D.-D. Lee, *Sens. Actuators, B*, 2001, **B77**, 228.
62 Q. Zhou, R. Minhas, A. Ahmad and T. A. Wheat, *Environ. Conscious Mater.: Ecomater., Proc. Int. Symp*., 2000, 521.
63 E. Llobet, J. Rubio, X. Vilanova, J. Brezmes, X. Correig, J. W. Gardner and E. L. Hines, *Sens. Actuators, B*, 2001, **B76**, 419.
64 M. Penza, G. Cassano, F. Tortorella and G. Zaccaria, *Sens. Actuators, B*, 2001, **B73**, 76.
65 D. S. Lee, H.-Y. Jung, J.-W. Lim, M. Lee, S.-W. Ban, J.-S. Huh and D.-D. Lee, *Sens. Actuators, B*, 2000, **B71**, 90.
66 J. J. Lavigne, S. Savoy, M. B. Clevenger, J. E. Ritchie, B. McDoniel, S.-J. Yoo, E. V. Anslyn, J. T. McDevitt, J. B. Shear and D. Neikirk, *J. Am. Chem. Soc*., 1998, **120**, 6429.
67 A. Goodey, J. J. Lavigne, S. M. Savoy, M. D. Rodriguez, T. Curey, A. Tsao, G. Simmons, J. Wright, S.-J. Yoo and Y. Sohn, *et al*., *J. Am. Chem. Soc*., 2001, **123**, 2559.
68 C. Krantz-Rulcker, M. Stenberg, F. Winquist and I. Lundstrom, *Anal. Chim. Acta*, 2001, **426**, 217.
69 A. Legin, A. Smirnova, A. Rudnitskaya, L. Lvova, E. Suglobova and Y. Vlasov, *Anal. Chim. Acta*, 1999, **385**, 131.
70 C. Di Natale, F. Davide, J. A. J. Brunink, A. D'Amico, Y. G. Vlasov, A. V. Legin and A. M. Rudnitskaya, *Sens. Actuators*, 1996, **B34**, 539.
71 J.-M. Lehn, *Supramolecular Chemistry: Concepts and Perspectives*, VCH, Weinheim, Germany, 1995.
72 *Comprehensive Supramolecular Chemistry*, ed. J. L. Atwood, J. E. D. Davies, D. D. MacNicol, F. Vogtle and J.-M. Lehn, Pergamon, Oxford, UK, 1996.
73 J.-M. Lehn, *Supramolecular Science: Where It Is and Where It Is Going*, ed. R. Ungaro and E. Dalcanale, Kluwer, Dordrecht, Netherlands, 1999, pp. 287–304.
74 J. J. Lavigne and E. V. Anslyn, *Angew. Chem., Int. Ed*., 2001, **40**, 3118.
75 M. Schaeferling, *Springer Ser. Chem. Sens. Biosens*., 2005, **3**, 45.
76 A. L. Kukla, A. S. Pavluchenko, Y. M. Shirshov, N. V. Konoshchuk and O. Y. Posudievsky, *Sens. Actuators, B*, 2009, **B135**, 541.
77 B. Li, S. Santhanam, L. Schultz, M. Jeffries-El, M. C. Iovu, G. Sauve, J. Cooper, R. Zhang, J. C. Revelli and A. G. Kusne, *et al*., *Sens. Actuators, B*, 2007, **B123**, 651.
78 M. J. Gismera, S. Arias, M. T. Sevilla and J. R. Procopio, *Electroanalysis*, 2009, **21**, 979.
79 D. P. A. Correia, J. M. C. S. Magalhaes and A. A. S. C. Machado, *Microchim. Acta*, 2008, **163**, 131.
80 H.-G. Jahnke, A. Rothermel, I. Sternberger, T. G. A. Mack, R. G. Kurz, O. Paenke, F. Striggow and A. A. Robitzki, *Lab Chip*, 2009, **9**, 1422.
81 A. L. Ghindilis, M. W. Smith, K. R. Schwarzkopf, C. Zhan, D. R. Evans, A. M. Baptista and H. M. Simon, *Electroanalysis*, 2009, **21**, 1459.
82 R. A. Potyrailo, C. Surman and W. G. Morris, *J. Comb. Chem*., 2009, **11**, 598.
83 K. Morita and Y. Shimizu, *Anal. Chem*., 1989, **61**, 159.
84 T. A. Dickinson, J. White, J. S. Kauer and D. R. Walt, *Nature*, 1996, **382**, 697.
85 M. Schena, in *Microarray Analysis*, Wiley-Liss, Hoboken, NJ, 2003, p. 630.
86 A. Goodey, J. J. Lavigne, S. M. Savoy, M. D. Rodriguez, T. Curey, A. Tsao, G. Simmons, J. Wright, S. J. Yoo and Y. Sohn, *et al*., *J. Am. Chem. Soc*., 2001, **123**, 2559.
87 N. A. Rakow and K. S. Suslick, *Nature*, 2000, **406**, 710.
88 M. Kitamura, S. H. Shabbir and E. V. Anslyn, *J. Org. Chem*., 2009, **74**, 4479.
89 A. W. Martinez, S. T. Phillips, E. Carrilho, S. W. Thomas, H. Sindi and G. M. Whitesides, *Anal. Chem*., 2008, **80**, 3699.
90 R. A. Potyrailo, W. G. Morris, A. M. Leach, T. M. Sivavec, M. B. Wisnudel and S. Boyette, *Anal. Chem*., 2006, **78**, 5893.
91 D. B. Hibbert, *Electroanalysis*, 1998, **10**, 1077.
92 S. M. Scott, D. James and Z. Ali, *Microchim. Acta*, 2006, **156**, 183.
93 K. R. Beebe, R. J. Pell and M. B. Seasholtz, *Chemometrics: A Practical Guide*, Wiley, New York, 1998.
94 M. Otto, *Chemometrics: Statistics and Computer Application in Analytical Chemistry*, Wiley-VCH, Weinheim, New York, 2007.
95 P. C. Jurs, G. A. Bakken and H. E. McClelland, *Chem. Rev*., 2000, **100**, 2649.
96 A. T. Wright and E. V. Anslyn, *Chem. Soc. Rev*., 2006, **35**, 14.
97 R. G. Brereton, *Chemometrics for pattern recognition*, Wiley, Chichester, UK, 2009.
98 R. G. Brereton, *Applied Chemometrics for Scientists*, Wiley, Chichester, 2007.
99 E. R. Malinowski, *Factor Analysis in Chemistry*, John Wiley & Sons, New York, 2nd edn, 1991.
100 M. Jambu, *Exploratory and Multivariate Data Analysis*, Academic Press, Boston, 1991.
101 B. A. Suslick, L. Feng and K. S. Suslick, *Anal. Chem*., 2010, **82**, 2067.
102 M. A. Palacios, R. Nishiyabu, M. Marquez and P. Anzenbacher, Jr., *J. Am. Chem. Soc*., 2007, **129**, 7538.
103 Z. Wang, M. A. Palacios and P. Anzenbacher, Jr., *Anal. Chem*., 2008, **80**, 7451.
104 M. A. Palacios, Z. Wang, V. A. Montes, G. V. Zyryanov and P. Anzenbacher, Jr., *J. Am. Chem. Soc*., 2008, **130**, 10307.
105 J.-S. Lee, J. W. Lee and Y.-T. Chang, *J. Comb. Chem*., 2007, **9**, 926.
106 J. W. Lee, J. S. Lee, M. Kang, A. I. Su and Y. T. Chang, *Chem.–Eur. J*., 2006, **12**, 5691.
107 J. F. Folmer-Andersen, M. Kitamura and E. V. Anslyn, *J. Am. Chem. Soc*., 2006, **128**, 5652.
108 J. Tan, H. Wang and X. Yan, *Anal. Chem*., 2009, **81**, 5273.
109 A. Schiller, B. Vilozny, R. A. Wessling and B. Singaram, *Anal. Chim. Acta*, 2008, **627**, 203.
110 A. Ponnu, N. Y. Edwards and E. V. Anslyn, *New J. Chem*., 2008, **32**, 848.
111 A. Sen, J. D. Albarella, J. R. Carey, P. Kim and W. B. McNamara, *Sens. Actuators, B*, 2008, **B134**, 234.
112 B. E. Collins and E. V. Anslyn, *Chem.–Eur. J*., 2007, **13**, 4700.
113 H. Zhou, L. Baldini, J. Hong, A. J. Wilson and A. D. Hamilton, *J. Am. Chem. Soc*., 2006, **128**, 2421.
114 A. T. Wright, N. Y. Edwards, E. V. Anslyn and J. T. McDevitt, *Angew. Chem., Int. Ed*., 2007, **46**, 8212.
115 C. J. Musto, S. H. Lim and K. S. Suslick, *Anal. Chem*., 2009, **81**, 6526.

 This journal is © The Royal Society of Chemistry 2010

116 C. Zhang and K. S. Suslick, *J. Agric. Food Chem.*, 2007, **55**, 237.
117 G. Zyryanov, M. A. Palacios and P. Anzenbacher, Jr., *Angew. Chem., Int. Ed.*, 2007, **46**, 7849.
118 W. Carey, K. Beebe and B. Kowalski, *Anal. Chem.*, 1986, **58**, 149.
119 F. Avila, D. Myers and C. J. Palmer, *J. Chemom.*, 1991, **5**, 455.
120 E. Green, M. J. Olah, T. Abramova, L. R. Williams, D. Stefanovic, T. Worgall and M. N. Stojanovic, *J. Am. Chem. Soc.*, 2006, **128**, 15278.
121 M. Maeder and Y.-M. Neuhold, *Practical Data Analysis in Chemistry*, Elsevier, Amsterdam, 2007.
122 I. E. Frank, Modern nonlinear regression methods (tutorial), Chemometrics Intell. Systems 27, 1995, pp. 1–19.
123 K. Varmuza and P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*, CRC Press-Taylor & Francis, Boca Raton, 2009.
124 J. H. Ward, Jr., *J. Am. Stat. Assoc.*, 1963, **58**, 236.
125 G. W. Milligan, *Psychometrika*, 1980, **45**, 325.
126 S. H. Lim, C. J. Musto, E. Park, W. Zhong and K. S. Suslick, *Org. Lett.*, 2008, **10**, 4405.
127 Y. Koshi, E. Nakata, H. Yamane and I. Hamachi, *J. Am. Chem. Soc.*, 2006, **128**, 10413.
128 J. H. Bang, S. H. Lim, E. Park and K. S. Suslick, *Langmuir*, 2008, **24**, 13168.
129 N. T. Greene and K. D. Shimizu, *J. Am. Chem. Soc.*, 2005, **127**, 5695.
130 A. Schiller, R. A. Wessling and B. Singarum, *Angew. Chem., Int. Ed.*, 2007, **46**, 6457.
131 T. L. Nelson, I. Tran, T. G. Ingallinera, M. S. Maynor and J. J. Lavigne, *Analyst*, 2007, **132**, 1024.
132 M. S. Maynor, T. L. Nelson, C. O'Sullivan and J. J. Lavigne, *Org. Lett.*, 2007, **9**, 3217.
133 A. D. Hughes, I. C. Glenn, A. D. Patrick, A. Ellington and E. V. Anslyn, *Chem.–Eur. J.*, 2008, **14**, 1822.
134 M. E. Germain and M. Knapp, *J. Am. Chem. Soc.*, 2008, **130**, 5422.
135 T. Zhang, N. Y. Edwards, M. Bonizzoni and E. V. Anslyn, *J. Am. Chem. Soc.*, 2009, **131**, 11976.
136 O. R. Miranda, C. You, R. Phillips, I.-B. Kim, P. S. Ghosh, U. H. F. Bunz and V. M. Rotello, *J. Am. Chem. Soc.*, 2007, **129**, 9856.
137 C. You, O. R. Miranda, B. Gider, P. S. Ghosh, I.-B. Kim, B. Erdogan, S. A. Krovi, U. H. F. Bunz and V. M. Rotello, *Nat. Nanotechnol.*, 2007, **2**, 318.
138 M. De, S. Rana, H. Akpinar, O. R. Miranda, R. R. Orvizo, U. H. F. Bunz and V. M. Rotello, *Nat. Chem.*, 2009, **1**, 461.
139 A. Bajaj, O. R. Miranda, I.-B. Kim, R. L. Phillips, D. J. Jerry, U. H. F. Bunz and V. M. Rotello, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10912.
140 R. L. Phillips, O. R. Miranda, C. You, V. M. Rotello and U. H. F. Bunz, *Angew. Chem., Int. Ed.*, 2008, **47**, 2590.
141 J. Zupan and J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley, New York, 1999.
142 J. A. Burns and G. M. Whitesides, *Chem. Rev.*, 1993, **93**, 2583.
143 P. Ciosek and W. Wroblewski, *Analyst*, 2007, **132**, 963.
144 D. R. Walt, *Chem. Soc. Rev.*, 2010, **39**, 38.
145 A. Buryak, A. Pozdnoukhov and K. Severin, *Chem. Commun.*, 2007, 2366.
146 A. Buryak, F. Zaubitzer, A. Pozdnoukhov and K. Severin, *J. Am. Chem. Soc.*, 2008, **130**, 11260.
147 S. R. Johnson, J. M. Sutter, H. L. Engelhardt, P. C. Jurs, J. White, J. S. Kauer, T. A. Dickinson and D. R. Walt, *Anal. Chem.*, 1997, **69**, 4641.
148 T. A. Dickinson, K. L. Michael, J. S. Kauer and D. R. Walt, *Anal. Chem.*, 1999, **71**, 2192.
149 J. M. Sutter and P. C. Jurs, *Anal. Chem.*, 1997, **69**, 856.
150 D. Mikami, T. Ohki, K. Yamaji, S. Ishihara, D. Citterio, M. Hagiwara and K. Suzuki, *Anal. Chem.*, 2004, **76**, 5726.
151 H. He, G. Xu, X. Ye and P. Wang, *Meas. Sci. Technol.*, 2003, **14**, 1040.
152 D. Leung and E. V. Anslyn, *J. Am. Chem. Soc.*, 2008, **130**, 12328.
153 L. Zhu, S. H. Shabbir and E. V. Anslyn, *Chem.–Eur. J.*, 2006, **13**, 99.
154 S. Winters-Hilt, A. Yelundur, C. McChesney and M. Landry, *BMC Bioinformatics*, 2006, **7**, no pages given.
155 W. S. Noble, *Nat. Biotechnol.*, 2006, **24**, 1565.
156 O. Ivanciuc, *Rev. Comput. Chem.*, 2007, **23**, 291.
157 R. G. Brereton and G. R. Loyd, *Analyst*, 2010, **135**, 230.
158 T. Mayr, C. Igel, G. Liebsch, I. Klimant and O. S. Wolfbeis, *Anal. Chem.*, 2003, **75**, 4389.
159 M. J. Aernecke, J. Guo, S. Sonkusale and D. R. Walt, *Anal. Chem.*, 2009, **81**, 5281.
160 R. Pei, A. Shen, M. J. Olah, D. Stefanovic, T. Worgall and M. N. Stojanovic, *Chem. Commun.*, 2009, 3193.
161 S. J. Dixon and R. G. Brereton, *Chemom. Intell. Lab. Syst.*, 2009, **95**, 1.
162 *Microsystem Engineering of Lab-on-a-Chip Devices*, ed. O. Geschke, H. Klank and P. Telleman, Wiley-VCH, Weinheim, 2004.
163 *Lab on a Chip Technology: Fabrication and Microfluidics*, ed. K. Herold and A. Rasooly, Caister Academic Press, Norfolk, 2009.
164 *Micro Instrumentation*, ed. M. V. Koch, K. M. VandenBussche and R. W. Chrisman, Wiley-VCH, Weinheim, 2007.
165 G. Jung, *Combinatorial Chemistry: Synthesis, Analysis, Screening*, Wiley-VCH, Weinheim, 1999.
166 *Combinatorial Chemistry and Technologies: Methods and Applications*, ed. G. Fassina and S. Miertus, CRC Press, Boca Raton, 2005.
167 G. A. Ozin, A. Arsenault and L. Cademari, *Nanochemistry: A Chemical Approach to Nanomaterials*, RSC Publishing, Cambridge, 2nd edn, 2009.
168 P. Anzenbacher, Jr. and M. A. Palacios, *Nat. Chem.*, 2009, **11**, 80.
169 G. K. Geiss, R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D. L. Dunaway, H. P. Fell, S. Ferree, R. D. George and T. Grogan, *et al.*, *Nat. Biotechnol.*, 2008, **26**, 317.